University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 275

Multigrading and Child Achievement

Gian Paolo Barbetta, Giuseppe Sorrenti and Gilberto Turati

Revised version, February 2019

Multigrading and Child Achievement*

Gian Paolo Barbetta[†] Giuseppe Sorrenti[‡] Gilberto Turati[§]

February 2019

Abstract

We study how grouping students of different grades into a single class (multigrading) affects children's cognitive achievement. To do so, we build instruments to identify the causal effect of multigrading by exploiting an Italian law that controls class size and grade composition. We focus on seven- and ten-year-old second and fifth graders, respectively. Results suggest that attendance in multigrade versus single-grade classes increases students' performance on standardized tests by 19 percent of a standard deviation (24 percent, gross of the class size effect) for second graders, while it has zero effect for fifth graders. The positive impact of multigrading only appears for children sharing their class with peers from higher grades and it is relatively stronger for students from disadvantaged backgrounds.

JEL classification: I26, I28, R53

Keywords: Multigrade classes, child development, peer effects, rural areas

---

*We wish to thank Gianluca Argentin, Massimiliano Bratti, Claudio Lucifora, and Ulf Zölitz for helpful comments on preliminary versions of this paper, Gianna Barberi and Lucia De Fabrizio (MIUR) for kindly providing data, and Patrizia Falzetti (INVALSI) for both data and comments on a preliminary draft. Chiara Paola Donegani provided excellent research assistance when this project was initiated several years ago. Financial support from the Swiss National Science Foundation (100018_165616) is gratefully acknowledged (Sorrenti).

[†]Università Cattolica del Sacro Cuore, Milan (IT). E-mail: gianpaolo.barbetta@unicatt.it

[‡]University of Zürich and Jacobs Center for Productive Youth Development (CH). E-mail: giuseppe.sorrenti@econ.uzh.ch

[§]Università Cattolica del Sacro Cuore, Rome (IT). E-mail: gilberto.turati@unicatt.it

1 Introduction

Education in childhood aims to foster cognitive skills of pupils as well as their individual talents and social traits (Cunha and Heckman, 2008, 2010). These skills are largely predictive of future individual success in school, employment, and life in general (Heckman et al., 2006; Agostinelli and Wiswall, 2016). Therefore, analyzing the effect of programs, practices, or policies that could interact with the development of individual skills during childhood—and early childhood in particular—is a priority (Cunha et al., 2010).

In this paper we study the effect of "multigrading," the practice of mixing more than one grade in a class, on early childhood cognitive development. We focus on seven-year-old (second grade) and ten-year-old (fifth grade) students in Italian primary school and implement an instrumental variable (IV) approach to address the endogeneity concerns (for example parental preferences) underlying attendance in multigrade classes.[1] We also separate the effect of class composition in terms of grades from the effect of class size.

Attendance in a multigrade class potentially affects child development via different channels. Multigrade classes favor interactions among peers who are either more, or less, mature; these interactions could directly foster or slow the acquisition of cognitive skills, including mathematical and linguistic abilities. At the same time, attending a class with peers of different ages is likely to influence—with unknown results—the noncognitive abilities of children, impacting their social and emotional development and influencing friendships and self-perception, as well as other behavioral traits such as altruism or attitudes toward schooling. Finally, teaching practices and methods might also be influenced by grade composition of the class, again affecting children's learning.

Our IV identification strategy takes advantage of Italian law DPR 81/2009, a law that regulates the creation of both single-grade and multigrade classes. This law prescribes cutoffs, in terms of students enrolled in a specific grade, which affect the number of classes for each grade in a specific school (class size) and grade composition of classes (single-grade versus

[1]Primary school in Italy covers first through fifth grades.

multigrade). These cutoffs define intervals in the number of enrolled students for each grade that work as instruments for the actual grade composition of classes and for class size. We exploit these intervals in a Maimonides' Rule fashion (see Angrist and Lavy, 1999; Angrist et al., 2017) to predict both the individual probability of being assigned to a multigrade class and the number of students in a given class.

Our outcome variable is represented by the performance of second and fifth graders on the INVALSI standardized test in math and language. The INVALSI test is given in all Italian primary schools by the National Institute for the Evaluation of the Instruction and Training System. The sample of our analysis focuses on students attending school in municipalities where no more than one primary school operates. This restriction depends on data limitations that make it impossible to identify multigrade classes in municipalities where more than one primary school is active. Nonetheless, this limit mitigates the endogeneity concerns related to school choice and its interaction with parental preferences. In fact, in our sample, parents can only choose one school for their children (the one located in the municipality where they live), unless they decide to bear the costs of commuting to a more distant school in a different municipality.

We find that multigrade attendance increases second-grade students' math and language standardized test scores by about 16–24 percent of a standard deviation; for fifth graders, our findings suggest a statistically insignificant effect of attending a multigrade class. As the large majority of fifth graders in a multigrade class have experienced several years of primary school in a multigrade class, we interpret the result for fifth graders as the cumulative effect of multigrading at the end of primary school. Results for second and fifth graders are robust to: (i) excluding class size from the regression or considering class size as a standard control versus an endogenous variable and (ii) a complete set of sensitivity checks.

The multigrade effect for second-grade students is heterogeneous with respect to children's characteristics such as gender and family socioeconomic status. Females benefit more from multigrading than their male counterparts. We also unveil that children from low socioeconomic backgrounds (proxied by parental education) obtain greater benefits from multigrading. The latter result suggests that multigrading could mitigate the effect of poor socioeconomic conditions on child development; these conditions have been identified as crucial ingredients for explaining differences in cognitive achievement (Todd and Wolpin, 2007).[2] There is no evidence of heterogeneous effects of multigrading for fifth graders.

We investigate the mechanism underlying the effect of multigrading on cognitive development. We start with the analysis of grade composition of multigrade classes attended by second-grade students. We find that the overall positive effect of multigrading on second graders is driven by children's sharing their class with students from higher grades. The interaction with more-mature peers and the exposure to a more refined vocabulary are likely to shape this positive multigrade effect. On the contrary, multigrading does not show beneficial effects on second graders attending a multigrade class with younger peers.

We next consider fifth-grade students and then, to better interpret the lack of effect of multigrading, we track back their entire primary school career. Three-quarters of students in a multigrade class in their fifth grade of primary school had been in a multigrade class for at least three of the preceding four grades. Hence, the high persistence of multigrading suggests that the multigrade impact on fifth graders needs to be interpreted as a cumulative multigrade effect at the end of primary school. Thanks to the career path reconstruction, we provide additional evidence on the multigrade experience at the beginning of primary school.[3] In the test administered at the end of fifth grade, students who were in a multigrade class exclusively in first or second grade overperform students who never attended a multigrade class. On the contrary, the experience of multigrading only in fifth grade, the last year of primary school, negatively impacts achievements.[4]

[2]Practices that help children from poor backgrounds are extremely relevant given that child poverty is a massive phenomenon worldwide. In Italy alone, around 1.3 million children (12.5 percent) were living in poverty in 2016 (ISTAT, 2017).

[3]Namely, the grades in which students in multigrade classes are highly likely to share the classroom with more-mature peers.

[4]Admittedly, due to the high persistence of multigrading, the sample sizes of students experiencing a multigrade class only at the beginning or only at the end of primary school is relatively small. Reduced sample sizes translate into a lower precision of our estimates.

Understanding the effect of multigrading on child development is an important topic. Multigrading is a widespread practice that accounts for about one-third of the total number of classes worldwide according to UNESCO (2004). It is particularly common in remote and less affluent areas of the world such as many areas in developing countries, where its widespread use is often driven by economic constraints, and where children are also more likely to be exposed to fewer learning opportunities. For example, 78 percent of schools in Peru were multigrade in 1998. In Sri Lanka or Vietnam, multigrading is the only available option for children living in poor or remote areas (Hargreaves et al., 2001).

Multigrading is a common practice in developed countries as well. For instance, 28 percent of schools in the United States reported using multiage grouping in 2007, and in France, 37 percent of primary school students attend a multigrade class (Leuven and Rønning, 2016). Multigrade classes account for 70 percent of the classes in Finland and 53 percent in the Netherlands (Mulkeen and Higgins, 2009)—countries whose students generally achieve excellent scores on international standardized tests such as the OECD PISA test. In Italy, absent official statistics and according to our analysis, about 20 percent of schools located in municipalities with no more than one primary school have adopted multigrading.

Multigrade classes have won approval among pedagogists and educational psychologists and have been advanced in several quarters. For instance, they are one of the key ingredients of Montessori schools, which mushroomed in developed countries in an effort to obtain better educational results.[5] Multigrade classes are often proposed as new pedagogical tools that can better adapt to each pupil's rhythm of learning. In Switzerland—starting in 2003 and for the subsequent seven years—the pilot "Basisstufe project" (literally, "Base level" project) grouped more than 3,000 four- to eight-year old students in the same class, for about 150

[5]Montessori schools—named after the Italian pedagogist Maria Montessori—consider the use of mixed-age classrooms essential for their educational approach (see, for example, the website of the Association Montessori Internationale). More generally, advocates of multigrade classes say that "the traditional approach of dividing students into single grades based on an arbitrary birth-date range is illogical. ... Multiage education ... puts learners at the center, socially and academically. On the social side, younger children look for guidance to older students who know the ropes, while the older students in the classroom organically learn about mentoring, leadership, and collaboration." (The Atlantic, May 9, 2017).

classes.[6] In Italy, the project "Piccole scuole" (literally, "Small schools"), started by the public National Institute for Innovation and Research in Education (INDIRE) in 2015, helps teachers of multigrade classes in remote areas to work together aided by Information and Communications Technology (ICT).

Despite the widespread use and support of multigrading, its effect on student achievement has been the topic of very few solid empirical studies. Evidence of the impact on very young children is even more scarce. Early studies—surveyed in Little (2001)—are unable to properly address sorting of students into multigrade classes. An exception is the work by Leuven and Rønning (2016), which studies how classroom grade composition affects fifteen-years-old students' achievements in Norwegian junior high schools. By exploiting a national regulation determining classroom grade composition, the authors show that a one-year exposure to a class that combines two grades increases performance by about 4 percent of a standard deviation. On the contrary, Checchi and De Paola (2018) analyze ten-years-old students in Italy and find a cumulative negative impact of multigrading on cognitive achievement; albeit, the effect is statistically significant only for numeracy. We extend many aspects of their multigrading evaluation. First, while they can only try to infer whether students were placed in a multigrade class, we develop an algorithm to precisely identify multigrade classes. This point is extremely important as a considerable fraction of classes that should be multigrade classes according to the law are indeed single-grade classes. The actual identification of multigrade classes allows us to properly disentangle the multigrade effect from the class size effect. Second, while they limit their analysis to fifth-grade students, we also consider second graders. Third, we adopt a less restrictive sample definition, which allows us to better define counterfactuals and explain the mechanism behind our findings. Checchi and De Paola (2018) consider only schools that have no more than one classroom for fifth-grade students and also exclude classes smaller/larger than the expected size for multigrade classes. We consider municipalities that host no more than one primary school.

[6]See www.swissinfo.ch/ita/societa/scuola-e-territorio-_pluriclassi–retaggio-del-passato-o-pedagogia-del-futuro-/33582238.

This rule allows us to observe different mixtures of grades for second- and fifth-grade students.

Our work makes several contributions to the existing literature. First, we study the causal impact of multigrading on early childhood development and separate its effect from the effect of class size. We focus on children at the beginning of their schooling career. This is particularly important given that investment in early childhood education generates the highest rate of return (Cunha et al., 2010). Second, we show evidence of heterogeneous effects of multigrading with respect to characteristics such as students' gender and socioeconomic background. According to our results, children from low socioeconomic backgrounds obtain higher benefits from multigrading. Therefore, practices such as multigrading may act as an effective tool for lowering initial inequalities among students. Third, we further unveil the mechanism underlying the multigrade effect. We show the importance of the age of peers in a multigrade class. Moreover, by reconstructing the career path of fifth graders, we provide an interpretation for the absence of a multigrade effect for fifth graders. The effect observed in the fifth grade of primary schooling should be interpreted as the cumulative effect of multigrading at the end of primary school. This interpretation is in line with results in Leuven and Rønning (2016) in which the authors show that the presence of lower-grade peers in a multigrade class is detrimental for achievement.

The remainder of the paper is structured as follows. Section 2 provides essential background information on Italian primary schools and the rules governing class formation. Section 3 describes how we created our data. Section 4 presents the identification strategy. Section 5 discusses the results. Section 6 investigates the mechanism underlying the multigrade effect. Section 7 concludes.

2 The Institutional Background

Primary school (ISCED 1) in Italy begins for children when they are six years old; it covers first through fifth grades. Primary education is compulsory, and its main purpose is to provide sound basic training in reading, writing, and mathematics, plus an elementary understanding of subjects such as geography, history, science, English language, drawing, and music.

Parents can enroll their children in one of the more than 15,000 public primary schools (mostly state-run institutions) or in one of the about 1,500 private schools that operate in the country (2014 census by the Italian Statistical Office, ISTAT). Public schools enroll more than 93 percent of the approximately 2.8 million students attending primary school.

No official statistics about multigrading are available. Anecdotal evidence shows that most Italian primary school students attend a single-grade class. Multigrading is relatively common in Italy: according to our analysis, about 20 percent of schools located in municipalities with no more than one primary school adopt multigrading.

The estimation of the causal impact of multigrading on individual performance is difficult because students' selection into those classes could be nonrandom. In theory, parents as well as school principals and teachers could have specific preferences and could therefore try to modify class composition. In practice, this problem should not be very relevant in Italy for two main reasons. First, in creating classes, school principals must follow the rules established by the DPR 81/2009. This law defines a set of thresholds to determine: (i) when a new single-grade class should be created given the number of students in a single cohort, and (ii) whether a multigrade class should be formed. Thresholds are expressed based on the number of students of the same grade enrolled in a specific primary school and influence both the probability of ending up in a multigrade class and the size of classes. The law specifically establishes that:

- single-grade classes consist of a minimum of 15 and a maximum of 26 students;
- multigrade classes consist of a minimum of 8 and a maximum of 18 students;
- in special cases such as isolated villages, small islands, and areas characterized by the presence of linguistic minorities, single-grade classes could be created with a minimum

of 10 students. Besides these special cases, the law allows some flexibility (reducing the maximum number of students per class) in the presence of disabled children.

Although the law is vague on this point, in practice, in the same school, students in a specific grade never attend a single-grade class while others are assigned to a multigrade class. Indeed, it would be difficult to justify assigning some students to a single-grade class while assigning others from the same cohort to a multigrade class.

Second, parental preferences are constrained by the specific enrollment process. In fact, public schools adopt uniform criteria to admit students, the main one being the distance between the student's house and the school. Students living in each school catchment area are automatically accepted, but students coming from outside the area can be accepted only if the school has spare capacity. Moreover, national rules require families to apply to a primary school by January–February each year, well before the beginning of the following school year (SY), which starts in mid-September. School principals are required to advise each family whether their children have been accepted within a month after their application. However, students are assigned to classes (and teachers are assigned to each class) only during the summer. Parents cannot participate in this process and they only learn of the class composition and the teachers' names shortly before the beginning of the SY (or even the first day of school).

The enrollment process characteristics play an important role in our identification strategy (Section 4) as they mitigate possible selection-into-schools endogeneity concerns. The institutional setting makes it very difficult for parents concerned about grade composition and class size to opt for alternative primary schools. Moreover, public school principals face constraints in exercising their possible preferences on this issue. However, although the effect of parents', school principals', and teachers' preferences on class composition appears to be relatively unimportant, we extensively address endogeneity concerns in the remainder of the paper.

3 The Data

Our aim is to compare the educational achievement of children attending either a single-grade or a multigrade class during primary school. We measure achievement using individual student scores on the national standardized test run by INVALSI. The INVALSI written test is intended to monitor the skills and knowledge of Italian students in two main areas, namely mathematics and language. Each test includes a set of multiple-choice items followed by open response questions. Students must conclude the tests in 45 to 90 minutes, depending on grade and subject.[7] The test was introduced in 2007 by law 176/2007, and it is administered yearly to second-, fifth-, eighth-, and tenth-grade students attending public or private schools.

We focus our analysis on second- and fifth-grade students, respectively, seven- and ten-year-old pupils. Although each school knows the individual scores of its students, public data about individual performance on the INVALSI test are fully anonymous: students, classes, and schools cannot be identified. This makes it impossible to detect the grade composition of each class using the INVALSI data alone. However, the INVALSI data contain some geographical and demographic information (e.g. province, population, size, and altitude of the municipality) that is fundamental for our matching procedure.

More precisely, we assembled a new data set that merges individual performance on the INVALSI test in the 2012/2013 SY with information included in two different administrative archives: i) School Register data[8] provided by the Italian Ministry of Education (MIUR), which contain detailed information about each Italian primary school, including the number of multigrade classes; and ii) Municipality Register data produced by ISTAT, which include the same geographical and demographic information for each Italian municipality as contained in the INVALSI data set. We use data about municipalities to bridge information in the INVALSI data and in the School Register data, and we create specific algorithms to

[7]More information about the INVALSI test is available at www.invalsi.it.

[8]This data set is built on administrative data coming from the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) - Rilevazione integrativa.

identify students attending a multigrade class.[9]

This procedure allows us to identify municipalities that host a single primary school, the name and the characteristics of this school, the educational achievement of its second- and fifth-grade students, the grade composition of their classes (single versus multigrade) as well as class sizes. As a result, our final data set includes the entire population of Italian second- and fifth-grade students attending a primary school located in a municipality hosting only one primary school. We end up with 4,295 primary schools out of 15,248 covered in the School Register data in the 2012/2013 SY, about 92,000 second-grade students and 90,000 fifth-grade students out of the 500,000 in each cohort all over the country.[10]

In Italy, around 65 percent of municipalities with primary schools have no more than one such school, which reflects the fact that 53 percent of municipalities in Italy are rural (or inner areas) according to the Ministry of Economic Development classification. These areas are usually far from service provision centers (see Materiali UVAL, 2014). In these rural contexts, mostly represented by small municipalities, multigrade classes are common.

On the one hand, our consideration of municipalities that host no more than one primary school represents a potential data limitation. On the other hand, this limit allows us to keep under control the problem of nonrandom assignment of students into classes with different grade compositions. In fact, in municipalities with no more than one primary school, parental choice about their children's school enrollment is automatically ruled out unless parents decide to take them to a different municipality and bear commuting costs, which increase directly with the distance from the closest alternative solution (a variable we control for in the empirical analysis). To improve comparability between the schools with and without multigrade classes even further, we also test our model on two different restricted samples based on cohort sizes: 30 students (two times the minimum number of students required to

[9] Appendix A.1 provides a detailed description of the data construction process.

[10] We drop the two regions of Valle d'Aosta and Trentino Alto Adige from our analysis because, in these areas, the administration of primary and secondary schools is assigned to regional (or provincial) authorities. Consequently, the Ministry of Education does not collect registration information for these areas. Students from these two regions only account for 1.4 percent of the total sample size of INVALSI test takers.

form a single-grade class) and 60 students.

Table 1 shows summary statistics for our samples. For second graders, the average score on the mathematics standardized test is around 19 points (out of 32, or about 59 percent of correct answers), while it is slightly higher (25 points out of 39, or about 64 percent of correct answers) for the language test. Around 6 percent of students attend a multigrade class, and class size is around 19 pupils per class. For fifth graders, the average performance on the mathematics standardized test is around 28 points (out of 47, or about 59 percent of correct answers), while it reaches 63 points (out of 82, or about 77 percent of correct answers) for the language test. Around 5 percent of students attend a multigrade class, and class size is around 19 pupils per class.

Besides age, children's characteristics are similar across the two samples of second and fifth graders. Samples are balanced in terms of gender (around 50 percent), and in both samples, around 11–12 percent of the children have migrant parents. In terms of socioeconomic background, we consider three different levels of parental education: completed university, completed high school, and a residual category for all those holding below a secondary education diploma. Similar patterns emerge when we compare fathers and mothers. Most children in our samples (more than 70 percent) come from families in which parents have at most an upper secondary education. The percentage of university graduates is always lower than 10 percent, while 17–18 percent of parents have an education below the high school level.

The bottom panel of Table 1 provides useful information about the geographical characteristics of the schools. The sample covers the entire Italian territory and all five macro-regions (NUTS 1) are represented. The Northwestern area is the most represented (44–46 percent), followed by the South (18–20 percent), the Northeast (17 percent), the Central area (11 percent), and the Islands (8 percent). Unsurprisingly, municipalities where the schools are located are relatively small, with fewer than 5,000 inhabitants on average.

4 The Identification Strategy

The identification of the causal impact of multigrading on child achievement is a challenging task because parental choices could drive the enrollment in multigrade classes. Although, as discussed in Section 2, this kind of concern should be relatively minor in the Italian context because of the process of class formation that totally excludes parents, we address the endogeneity issue by implementing an instrumental variable (IV) identification strategy. Our IV strategy builds on the research design in Angrist and Lavy (1999), which is often referred to as the Maimonides' Rule.[11] In this work, the authors exploit class size cutoffs imposed by a rule in Israel to estimate the impact of class size on scholastic achievement. A similar strategy is also used by Leuven and Rønning (2016), who exploit institutional features significantly affecting grade composition in Norway to specifically estimate the impact of grade mixing on student achievement.

Estimating the multigrade effect is complicated by the possible impact of class size on student achievement. Class size is potentially correlated with the probability of attending multigrade classes and, at the same time, it can independently affect a child's learning process. To account for possible class size effects, we implement a triple approach. First, we consider a model that excludes class size from the explanatory variables of student achievement. Second, we include class size as a control variable in our baseline specification. Third, as class size might suffer from the same sources of endogeneity of multigrading, we replicate our approach treating it as an extra endogenous variable in the model.

The identification strategy of this study is based on DPR 81/2009, a law that defines a set of rules based on exogenous cutoffs to establish whether a new single or multigrade class should be created. Specifically, we use predicted-by-the-law grade composition of classes as well as predicted-by-the-law class sizes to instrument the actual grade composition of classes and class sizes, respectively. The Italian law defines different thresholds in terms

[11] Estimation strategies inspired by Maimonides-style rules are common in the literature about class size and class composition. Some examples of works based on similar concepts are Hoxby (2000), Gary-Bobo and Mahjoub (2013), Bonesrønning (2003), Leuven et al. (2008), and Dobbelsteen et al. (2002).

of the number of students of the same grade enrolled in a specific school. Single-grade classes should include a minimum of 15 and a maximum of 26 students; on the other hand, multigrade classes should include no fewer than 8 and no more than 18 students. In special cases such as isolated villages, small islands, and areas characterized by the presence of linguistic minorities, deviation from the rules is possible: classes can be created with a lower number of students. However, this number cannot be lower than ten students for single-grade classes.

Despite the flexibility of accommodating local requests in specific years, DPR 81/2009 identifies four different relevant intervals (based on the number of students enrolled in a specific grade) that strongly affect class size as well as the individual probability of being enrolled in a multigrade class. The first interval pertains to schools with fewer than 10 students in one grade. In this case, no single-grade class should be created, and all students should be assigned to a multigrade class.[12] The second interval covers schools with 10 to 14 students in one grade. In this interval, following school characteristics such as geographical localization, both a single or a multigrade class could be created. The third interval covers schools with 15 to 26 students. In this case, the probability of being enrolled in a multigrade class should be close to zero. The same applies for the last interval, namely, schools with more than 26 students in one grade. The number of students is too high to create a multigrade class; therefore, according to the law, more than one single-grade class should be created.[13]

These four intervals allow us to create four different instruments (indicators for cohort sizes within a specific interval in terms of number of students) that address the endogeneity of both multigrading and class size. Indeed, DPR 81/2009 identifies four different margins along the cohort size distribution: the first two margins (fewer than 10 and between 10 and 14 students) affect the individual probability of ending up in a multigrade class, while the

[12] Although the law in principle prevents the creation of single-grade classes with fewer than 10 students, in practice, around 20–25 percent of students enrolled in schools with fewer than 10 students in their grade attend a single-grade class. The variability in grade composition for very small schools is important for our analysis as it allows us to separate the class size effect from the effect of grade composition.

[13] Remember that it is never observed that in the same school some students enrolled in a specific grade are assigned to a multigrade class while others are assigned to a single-grade class.

last two margins (between 15 and 26 and more than 26 students) largely affect class size but have limited or no effect on multigrading. We will provide an in-depth discussion of these margins in Section 5.1.

The validity of our instrumental approach relies on different assumptions. To avoid violating the exclusion restriction, we need our instruments to only affect students' test scores via grade composition (single versus multigrade class) and class size. As already discussed, the exact number of students enrolled in a specific grade is unpredictable as in Italy each family is free to enroll children in every school nationwide, although students living in school's catchment area have priority. Moreover, the enrollment procedure and its timing make it particularly difficult (if not impossible) for parents to form reliable expectations about the probability of their child ending up in a class with specific characteristics in terms of size and grade composition. A second important assumption underlying the literature based on Maimonides-style IV approaches is the absence of ad hoc manipulation around cutoffs. As shown in Section 5.1, we do not find evidence of such manipulation.

Under these assumptions we define the following reference model:[14]

$$TestScore_{isj} = \beta_1 Multigrade_{isj} + \beta_2 ClassSize_{isj} + \boldsymbol{X'_{isj}\beta_3} + \alpha_j + u_{isj} \quad , \tag{1}$$

where i, s, and j stand for student, school, and geographical macro-area, respectively.[15] $TestScore$ is the student's performance on the standardized national INVALSI test. We will focus on test scores at the end of the second and fifth grades of primary school. Specifically, the outcome variable is the combination of math and language INVALSI standardized test scores. After normalizing both test scores with a mean of zero and a standard deviation of one, we create a combined score in math and reading, taking the average of the normalized reading and math scores. We then normalize the combined score.[16] $Multigrade$ is a dummy

[14]In one specification, we exclude class size from the list of explanatory variables to estimate the raw effect of multigrading.

[15]Remember that we only observe municipalities with a single primary school.

[16]The procedure is similar to the one used in Dahl and Lochner (2012) and Agostinelli and Sorrenti (2018).

variable that takes the value of one if the student is enrolled in a multigrade class, and $ClassSize$ represents the student's class size. The vector $\mathbf{X}$ contains (observable) factors likely to affect test scores. Specifically, we control for child characteristics such as age, gender, and nationality by distinguishing Italian nationals of Italy, first-generation migrants, and second-generation migrants. We proxy parental background by including in the model both the father's and mother's education (university graduate versus high school graduate versus other) and profession.[17]

The vector $\mathbf{X}$ also includes information about the population and the altitude of the municipality hosting the school, and the minimum car travel time needed to reach the closest alternative primary school from the school each student actually attends. Travel time to the closest school indicates the presence of alternative school options. If alternative primary schools are available, parents who dislike multigrade classes and who expect their child to end up in such a class might decide to enroll their child in the closest school offering single-grade classes. For this reason, we include this measure of travel distance in all our models. As a robustness check we also use this variable as an additional instrument for the individual probability of being enrolled in a multigrade class. This strategy should deal with the possible residual endogeneity not corrected by our standard IV approach based on the thresholds defined by the law disciplining class formation. Finally, to consider geographical differences across the country, we also include in our model a set of macro-region fixed effects (α_j) that capture the average effect on test score for regions in the Northwest, the Northeast, the Central area, the South, and the Islands.[18]

According to thresholds defined by the law, we define four instrumental indicator variables $I[CohortSize]$ based on intervals in the number of students enrolled (cohort size) in a specific grade. The first variable indicates schools with fewer than 10 students, the second variable indicates schools with between 10 and 14 students, the third indicates schools with between

[17] Unfortunately, the INVALSI data do not contain information about family income. However, educational level and profession for both parents represent good proxies.

[18] We test for different definitions of geographical areas in Section 5.3.

15 and 26 students, and finally, the fourth variable indicates schools with more than 26 students in the cohort. We use the cohort size of second graders for the analysis of second-grade student test scores, and we use the cohort size of fifth graders for fifth-grade student test scores.

We estimate two different IV specifications defining two different sets of first stages. In one specification, we only instrument $Multigrade$, while we either exclude $ClassSize$ to estimate the raw effect of being in a multigrade class or use it as a standard control variable. The first stage in this case is defined as:[19]

$$Multigrade_{isj} = \boldsymbol{I[CohortSize]'_s\gamma_1} + \gamma_2 ClassSize_{isj} + \boldsymbol{X'_{isj}\gamma_3} + \alpha_j + \epsilon_{isj} \quad . \tag{3}$$

In the model including $ClassSize$ as an additional endogenous variable, we replicate the same first stage as in Equation 2 and then add a first stage of the following form:

$$ClassSize_{isj} = \boldsymbol{I[CohortSize]'_s\delta_1} + \boldsymbol{X'_{isj}\delta_2} + \alpha_j + \varepsilon_{isj} \quad . \tag{4}$$

Because of possible serial correlation of the error term at the school level, all the models are estimated with standard errors clustered at the school level.[20]

[19]For the case in which class size is excluded from the list of explanatory variables, the first stage would take the following form:

$$Multigrade_{isj} = \boldsymbol{I[CohortSize]'_s}\gamma_1 + \boldsymbol{X'_{isj}\gamma_2} + \alpha_j + \epsilon_{isj} \quad . \tag{2}$$

[20]Notice that we are considering municipalities with just one school in the full sample; hence, clustering at the school level is equivalent to clustering standard errors at the municipal level.

5 The Effect of Multigrading on Child Achievement

5.1 First-Stage Estimates

In this section we introduce the first-stage estimates. A typical concern related to the adoption of Maimonides-style rules is that such an identification strategy conveys possible ad hoc manipulation around the cutoff to prevent the enforcement of specific class or grade compositions. We deal with this concern by comparing observable individual characteristics around the three relevant cutoffs (10, 15, and 27 students) for both second and fifth graders. Table 2 reports the analysis of a relevant set of children's individual characteristics (age, gender, nationality) and family characteristics (parental education) for the case of second graders. Table 3 replicates the analysis for fifth-grade students. We impose a two-student-interval around each cutoff, comparing, for instance, schools with 8 or 9 students with schools with 10 or 11 students. All the average values are remarkably similar around the cutoffs. The p-values for the differences in means in column (4) of both tables suggest the lack of manipulation by school principals around the two cutoff points at 10 and 15 students. We find similar results for the 27-student-cutoff: although some differences in means appear statistically significant for the sample of second graders, average values are remarkably close.

Having addressed the manipulation concern, we rationalize our instrumental variable approach in Figure 1. For each grade of interest, we provide a graphical representation of the relation between the instruments and the two endogenous variables of the model, namely grade composition and class size. The figure in each panel contains bins representing the average y-value for each ventile of the distribution of the number of enrolled students. Vertical lines display the cutoffs in terms of enrolled students that define the intervals used as instruments, while the solid horizontal lines represent the average y-value for each interval.

We start with the relation between the probability of attending a multigrade class as a function of the number of students enrolled in a certain school. We estimate the first stage as explained in Equation 3: for each student, we compute the predicted probability of being assigned to a multigrade class based on the set of her individual characteristics. Panels (a) and (c) report the predicted probability of ending up in a multigrade class for second and fifth graders, respectively. Both panels convey at least two important findings. First, students in schools with fewer than 10 second graders have a very high predicted probability (around 80 percent) of ending up in a multigrade class. The probability is lower (15–19 percent) for the 10-through-14-students interval. The probability drops to zero for the last two intervals identified by the law. Second, the graph confirms that not all the intervals identified by the law affect an individual's probability of ending up in a multigrade class.

Panels (b) and (d) depict the case of class size. The first cutoff imposed by the law does not affect class size. For both second- and fifth-grade students, there is no relevant difference in average class size around the 10-student cutoff. On the contrary, for the following cutoffs, the relation between class size and the rules defined by the law becomes clear: average class size increases in the cohort size interval of 15–26 students at which point then school principals are required to form two classes; therefore, class size drops after the 27-student cutoff.

This graphical analysis highlights that the law identifies two different sets of instruments. More precisely, the first two intervals (instruments) in terms of enrolled students in a specific grade help identify the impact of multigrading, while the two following intervals (instruments) help identify the impact of class size.

Tables 4 and 5 show the first-stage estimates for second- and fifth-grade students, respectively. In the two tables, columns (1) and (2) report results for multigrading as the only endogenous variable excluding or including class size as a standard control, while columns (3) and (4) report results for the model considering class size as a further endogenous variable. In the latter case, column (3) reports first-stage results for multigrading, while column (4) shows first-stage results for class size. All the tests for under- and weak identification suggest that the first stage is very precise, and the instruments are extremely relevant.

The effect of the four instruments on the probability of ending up in a multigrade class is similar across specifications. Being enrolled in a school with a number of enrolled students falling in the first two intervals (fewer than 10 students, 10–14 students) highly predicts the probability of ending up in a multigrade class. With respect to the omitted category (schools with more than 26 second graders), students in schools with fewer than 10 second graders increase their probability of being assigned to a multigrade class by more than 80 percentage points. This result is hardly surprising as the law forces the adoption of multigrade classes for these specific cases. The coefficient remains significant, but with a lower magnitude (0.10–0.19), for schools with 10–14 students in the relevant cohort. On the contrary, schools with 15–26 students display zero effect on the probability of being assigned to a multigrade class.[21]

Results in column (4) of both tables confirm the validity of our instruments in predicting class size. The first two intervals for enrolled students report similar coefficients with respect to schools with more than 26 students: on average, class size decreases by around five students for schools with fewer than 10 students or 10-to-14 students. Schools with 15-to-26 students have, on average, an extra student per class compared with schools with more than 26 students. These results pinpoint how cohort size affects observed class size. However, this effect, as already suggested by the graphical evidence in Figure 1, is invariant when we consider the two first intervals in terms of the number of students.

As for the role of other variables, the analysis of the first stage is important for unveiling the role of parental education in shaping individual probability of attending a multigrade class. Students of parents reporting a university degree display a zero and (in most of the cases) statistically insignificant increase in the likelihood of being enrolled in a multigrade class as opposed to students of parents with at most a high school diploma. This finding confirms that parents are unlikely to understand the individual probability of their children's being assigned to a multigrade class or, alternatively, that their background does not

[21] It is important to note that, although quantitatively very close to zero, the coefficient for the 15-to-26 students interval becomes statistically significant in the model that includes class size as a control variable (column 2).

systematically shape their preferences on this matter.

5.2 Second-Stage Estimates

This section presents the main results concerning the effect of multigrading on student achievement. We start with the analysis of reduced-form effects of our instruments on the measure for child cognitive achievement, namely, the combined math-language test score. In Figure 2, we visually represent the relation between test scores and the number of enrolled second (Panel (a)) and fifth graders (Panel (b)) in each school. Each figure contains bins representing the average test score for each ventile of the distribution of the number of students enrolled. Vertical lines display the cutoffs defining intervals used as instruments, while the solid horizontal lines represent the average test score for each interval.

The first two intervals are suggestive of the reduced-form effect on test score induced by attendance in a multigrade class. Class size is unaffected by this cutoff. The evidence for second graders suggests a sizable positive impact of multigrading. When we consider the cutoff most related to the probability of ending up in a multigrade class (less or more than 10 enrolled students), the average positive value of the cognitive score for students below the cutoff is almost double than the score for students in the following interval (10-14 students). When we consider the second cutoff (15 students), we find similar evidence, albeit with a smaller between-interval difference. Finally, the third cutoff depicts the reduced-form effects of class size on performance. Class size displays a negative effect on student performance: the average test score for students from schools with more than 26 second graders is lower than the one for students from smaller schools.

The case of fifth graders conveys a different message. For fifth graders, the effect of multigrading on scholastic performance seems close to zero. Some evidence of the negative class size effect arises, although it is less evident than for second graders. In Appendix Table A.1, we provide the reduced-form estimates of our model. All the findings are in line with the graphical representation and hint at a likely positive effect of multigrading for second-grade students and at a possible negligible effect for fifth-grade students.

Table 6 shows second-stage estimates of the model in Equation 1 for the samples of second graders (columns 1-4) and fifth graders (columns 5-8). For each sample, we estimate the reference OLS model (columns 1 and 5), the IV model providing the raw effect of multigrading without controlling for class size (columns 2 and 6), the IV model with class size as a control variable (columns 3 and 7), and the IV model with class size as an endogenous variable (columns 4 and 8).

Multigrading displays a positive effect on the performance of second-grade students. In the OLS framework, attendance in a multigrade class increases, by as much as 9 percent of a standard deviation, the combined math-language test score. The IV analysis suggests a 24 percent of a standard deviation increase in the specification without the control for class size. As this coefficient is the combination of the multigrade effect and the class size effect, it is important to mention that multigrade classes have, on average, five to six fewer students than single-grade classes (Table 6). The effect of multigrading on test scores reaches the value of 16-19 percent of a standard deviation when we include class size as a standard control variable or as an additional endogenous variable in the model.

As for the performance of fifth graders, we do not find significant effects of multigrading. All specifications provide coefficients remarkably close to zero (either positive or negative) and never statistically significant at the usual confidence levels. The zero effect for fifth graders may be the result of: (i) students who were assigned to a multigrade class only in the fifth year of primary school; or (ii) students who attended more than one year in a multigrade class, so that their performance in the fifth year of primary school would represent the cumulative effect of attending a multigrade class. The first case is particularly rare. In the second case, the multigrade effect would sum up different important dimensions such as grade compositions of multigrade classes and age of peers. As will be discussed in Section 6.2, most students who attend a multigrade class in fifth grade attended multigrade classes in the previous years. Around 88 percent of students in a multigrade fifth grade class attended at least three years of multigrading out of five. Because of this evidence and the additional results in Section 6.2, we interpret the zero effect of multigrading for fifth graders as the cumulative effect of multigrading measured at the end of primary school.

The class size effect is remarkably similar when we consider different specifications and samples. For both second and fifth graders, class size plays a significant role in affecting achievement: a one-student-per-class increase explains an average decrease in individual performance of around one percent of a standard deviation. It is important to note that point estimates for the effect of multigrading are almost unaffected by the inclusion of class size as a pure control or as an endogenous variable.

Second-stage estimates require additional discussion. First, it is important to compare the OLS results with the ones obtained through IV. For second graders, although OLS and IV provide the same qualitative evidence on the effect of multigrading, the IV coefficients are higher in magnitude. Different factors may explain this difference such as omitted variable bias in OLS estimates, unobservable selection processes, and measurement error due to possible (reporting) errors in the administrative data we use to identify multigrade classes.

Second, we have extensively shown how the four instruments identified by the law differently affect the two endogenous variables of our model, namely multigrading and class size. According to first-stage estimates in Tables 4 and 5, the first two intervals determine the probability of ending up in a multigrade class, while the third and fourth intervals mainly affect class size. We therefore re-estimate our baseline model considering both $Multigrade$ and $ClassSize$ as endogenous variables, but using two different sets of instruments for each of the two variables. The two sets are the ones suggested by the graphical evidence and the previously shown first-stage estimates. Specifically, to instrument multigrade, we exploit the first three intervals in terms of enrolled students (fewer than 10 students, 10–14 students, and more than 14 students), while we consider the intervals of 1 to 14 students, 15 to 26 students, and more than 26 students as instruments for class size. Table 7 reports results for this exercise. First-stage estimates and diagnostic tests for both endogenous variables testify to a very precise first stage. Results are remarkably similar to those in the baseline models for the second stage. We record a positive impact of multigrading on the combined math-language test score for second-grade students, while the effect for fifth graders is statistically insignificant. Larger class size negatively affects the performance of both second and fifth graders.

Third, we further discuss the possible existence of parents' preference for single versus multigrade classes. Such preference is a potential source of endogeneity underlying individual enrollment in a multigrade class. First-stage evidence (see Section 5.1) regarding the role of parental education in shaping multigrade class attendance signals that this is unlikely to be a threat to the reliability of our findings. However, to be even more cautious, we estimate an additional IV specification in which travel time to the closest school is also used as an instrument. Assuming that parental preferences about grade composition play a role in choosing a school for their children, we have to consider that these preferences are constrained by the time needed to reach the closest alternative school. First-stage estimates suggest that travel time to the closest school plays a modest role in determining the individual probability of being enrolled in a multigrade class.[22] For second-grade students, an additional 1-minute distance causes an increase in the probability of enrollment in a multigrade class of 0.1 percentage points. The coefficient is statistically significant at the five percent level. The effect is statistically insignificant for fifth-grade students. No effect of travel time on class size is detected. Table 8 reports second-stage estimates. Main results are unaffected by the inclusion of travel time to the closest school as an additional instrument. Similar to our baseline analysis, the effect of multigrading on child cognitive achievement ranges between 16 and 19 percent of a standard deviation for second graders, while the effect for fifth graders is close to zero.

Finally, an additional source of concern comes from differences in school size when comparing schools with multigrade classes with schools with single-grade classes. This concern

[22] Appendix Table A.2 shows the first-stage estimates of the model.

is mitigated by the evidence that a considerable fraction of students in schools with fewer than ten students enrolled attend a single-grade class instead of the predicted-by-the-law multigrade class. Although the law, in principle, prevents the creation of single-grade classes with fewer than ten students, in practice they exist. In our sample, about 25 percent of students enrolled in schools with fewer than ten second graders attend a single-grade class. In the case of fifth graders, about 21 percent of students enrolled in schools with fewer than ten fifth graders attend a single-grade class. This evidence, on the one hand, makes the use of an instrumental variable approach essential to deal with possible endogeneity. On the other hand, single-grade classes with fewer than ten students are crucial for our analysis as they allow us to separately estimate the class size effect and the effect of grade composition. However, to be even more cautious about the definition of our (single-grade) control group, we re-estimate the baseline specifications on two different subsamples. Specifically, we limit the analysis to schools with a cohort size of second or fifth graders of no more than 60 or 30 students. The choice is driven by the law, which fixes the minimum number of students needed to form a single-grade class at 15. IV estimates are reported in Table A.3. All the main conclusions hold on these two different subsamples: the impact of multigrading for second-grade students is bounded between 15 and 21 percent of a standard deviation; for fifth-grade students, the impact is close to zero and statistically insignificant. Class size displays the same results as in previous analysis.

5.3 Sensitivity Tests

In this section we test the sensitivity of our results to the use of different definitions of the variables or the inclusion of different control variables. In the baseline analysis, we have considered the combined math-language test score as the main outcome of interest. We focus here on its two components separately. Appendix Table A.4 shows the effect of multigrading classes on each single test score, namely mathematics and language. The effect for second-grade students is positive and statistically significant for both outcomes. This guarantees that the overall effect shown in our baseline analysis is not exclusively driven by one single subject. The effect on mathematics scores seems slightly higher in magnitude (17–20 percent of a standard deviation) compared to the effect on language scores (12–14 percent of a standard deviation), although the difference between the two is not statistically significant. There is no significant effect on test scores for fifth graders.

In Appendix Table A.5, we test different models' specifications. In the first test (Panel (a)) we tackle one of the limitations of the INVALSI data: missing information about parents or about some of their characteristics, such as education or job. Although in our baseline analysis we introduced residual groups for students with missing parental information, here we restrict our sample to include only students for whom information about both parents' profession and educational level are available.[23] Results are remarkably similar to those in the baseline analysis.

In Panels (b) and (c) we investigate the possible geographical connotation of the multi-grading effect. This analysis is important per se, as it allows to infer possible heterogeneity at the local level. At the same time, it allows us to deal with some of the concerns related to possible bias induced by cheating and opportunistic behavior on the INVALSI test. We will discuss this point in detail below. In our main analysis, we use the five macro-regions (NUTS 1) to capture macro-regional fixed effects. Here we estimate two alternative models by considering regional fixed effects (NUTS 2, Panel (b)) and provincial fixed effects (NUTS 3, Panel (c)). Results remain unchanged in both specifications, suggesting that the choice of geographical level of aggregation does not affect the size and significance of our findings.

Finally, we discuss concerns related to possible opportunistic behavior on standardized tests. The use of scores on standardized tests to assess individuals' skills is common in social sciences such as economics, sociology, and psychology. However, given that standardized tests are useful tools to compare different schools, classes, and teachers, this produces potential incentives for opportunistic behavior by principals, teachers, and even students. For this

[23]We keep these observations in the baseline model as we want to also consider single-head households in our analysis.

reason, many scholars find these tests unreliable by providing growing evidence of cheating and score manipulation.[24] Bertoni et al. (2013) provide the first empirical evidence of the possible existence of cheating behavior on the INVALSI test. According to Lucifora and Tonello (2015), cheating mainly occurs when teachers shirk or decrease monitoring. In analyzing class size and score manipulation in the Southern regions, Angrist et al. (2017) find that cheating largely reflects teacher behavior, motivated by moral hazard in their grading effort. According to their estimates, roughly 5 percent of Italian scores are biased because of cheating.

Although it is impossible to conclusively identify cheating, here we show evidence about its potential impact on our estimates.[25] Several factors suggest that cheating should not be a major concern for our analysis. First, the existence of cheating in our setting would imply that our outcome of interest is the real score of the test plus some noise. On the one hand, if noise is stochastic, this would only affect our estimates by lowering precision, and all the coefficients would remain consistently estimated. On the other hand, if noise is correlated with our variable of interest (being enrolled in a multigrade class) the multigrade coefficient estimate would potentially be biased. As also confirmed by our discussions with administrators, principals, primary school teachers, and members of INVALSI, it is difficult to believe that the probability of observing opportunistic and cheating behavior directly depends on considering a single versus a multigrade class. Other elements suggested by the literature, such as teachers' unobserved characteristics, should be considered as the main determinants of possible cheating (Angrist et al., 2017). This anecdotal evidence is also confirmed by intraschool variability in cheating patterns.

[24]For instance, Jacob and Levitt (2003) estimate that, in Chicago public schools, serious cases of cheating by teachers or administrators occur in at least 4–5 percent of elementary school classrooms. Outside the United States, the debate about test score reliability has been raised in many countries such as the United Kingdom, Israel, France, and Sweden (e.g. Diamond and Persson, 2016).

[25]INVALSI provides a deterministic correction measure to address opportunistic behavior. This correction is based on a fixed predetermined rule considering, among other things, intraclass variance in scores. As confirmed by INVALSI, the correction is inappropriate for multigrade classes analyzed here. Schools with multigrade classes, as well as a considerable fraction of schools in our sample, are almost all, by definition, small schools.

We empirically deal with possible cheating-induced bias in our estimates with the analysis of geographical patterns underlying our baseline results. As shown by Bertoni et al. (2013), Angrist et al. (2017), and other qualitative studies, cheating is a major concern in southern Italy and much less so in northern regions.[26] With this evidence in mind, in Table 9, we replicate our baseline analysis focusing on regions in Northern Italy. Despite the reduced sample size, results for the Northern region are similar to the ones we obtain for the whole country. We observe a positive and significant effect of multigrading only for second-grade students (9–15 percent of a standard deviation), while for fifth graders, the impact is statistically insignificant. The coefficients of our relevant covariates in the sample from the Northern regions are never statistically different from those in the sample that includes the whole set of Italian regions.

5.4 Heterogeneous Effects of Multigrading

Our baseline analysis shows that classes mixing pupils of different ages are beneficial in terms of cognitive development for second graders (seven-year-old children), while there is no evidence of an effect for fifth graders (ten-year-old children). Are these effects the same for all children? We propose here a simple heterogeneity analysis based on two important dimensions: gender and family background.[27] Undeniably, the analysis of heterogeneous effects is always complicated in an IV setting. Instruments usually locally affect a fraction of the population, making it very difficult to compare different subpopulations. In our setting, the use of a law based on simple numerical rules to assign students to a single or multigrade class makes the analysis easier. Indeed, the instruments are also extremely powerful and

[26] According to the score manipulation index elaborated by Angrist et al. (2017), cheating only accounts for 2 percent of scores in the North and Central regions of Italy; this percentage is even lower for northern regions only. Appendix Figure A.1 (right panel) shows the geographical distribution at the provincial level of score manipulation in Angrist et al. (2017). Almost all the provinces in the northern part of the country are characterized by zero score manipulation. Ferrer-Esteban (2012) computes a similar measure for cheating by focusing on the exact repetition at the class level of the same sequence of answers (left panel of the Appendix Figure A.1)

[27] In Section 6 we focus on the effect induced by age composition of the class to understand the differential effect of studying with younger or more-mature peers. This analysis would also rationalize the interpretation of the underlying effect driving the zero impact of multigrading on fifth graders.

relevant when using subsamples.

Table 10 shows the analysis by gender (columns 1–2) and by parental education (columns 3–4). The upper panel focuses on second-grade students, while the bottom panel reports results for fifth graders. Scholastic performance is typically different when males and females are compared, with males usually underperforming in most subjects except mathematics. Results by gender suggest that, although differences across genders are not striking, second-grade females seem to benefit from multigrading more than second-grade males. The coefficient for multigrading increases by almost 30 percent (from 17 to 22 percent of a standard deviation) when females are compared to males. In general, both genders seem to benefit from multigrading, with girls obtaining higher benefits compared to boys. Results by gender for fifth graders confirm a statistically insignificant impact of multigrading.

In terms of parental background, we divide children in two groups according to their parents' education.[28] The low-background group (No one with university) includes children whose parents do not hold a university degree. The high-background group (One with university) includes children with at least one parent holding a university degree. As for second-grade students, multigrading positively shapes children's cognitive achievement for both parental backgrounds. However, the effect seems to be mainly driven by children from the lower parental background. The coefficient for the lower background is twice as large as the one for the higher background (20 versus 9 percent of a standard deviation). Moreover, the coefficient for children with at least one parent with a university degree is statistically insignificant.[29] As for fifth graders, the analysis according to parents' education confirms the zero effect stemming from multigrade.

The analysis of parental background highlights that very young children from less stimulating parental environments obtain the highest benefit from attendance in a multigrade class. This result identifies grade composition as a potential tool to mitigate the long-term

[28]Parental education is widely acknowledged as a good proxy for family socioeconomic background.

[29]It should be noted that in this specific case, the sample size for the higher parental background is considerably smaller than for the lower parental background. However, the instrument is relevant and strong in this subsample as well, making the concern of different sample sizes across the two groups less relevant.

effects of pupils' lower socioeconomic backgrounds. A class environment consisting of peers of different ages and, in particular, older peers, might act as an important additional input in the child development production function, partially compensating for the negative impact of low socioeconomic conditions.

6 Explaining the Multigrade Effect

In this section we shed light on the potential mechanism underlying the effects shown in the baseline analysis. We start by focusing on second-grade students. We then move to fifth graders and to the possible interpretations for the zero effect of multigrading on children's cognitive achievement at this specific stage of primary school.

6.1 The Case of Second Graders: Younger is Better?

To obtain a better understanding of the (positive) effect induced by multigrading during second grade on children's cognitive achievement, we replicate the baseline analysis by splitting the sample of second-grade students who attended a multigrade class into two different groups: students with only younger peers in the class and students with only older peers in the class.[30]

Table 11 illustrates results for second graders by grade composition of their class. Columns (1-3) show the analysis for multigrade classes in which the students share the classroom with younger peers. Columns (4-6) show the analysis for those students with older peers in their classroom. Columns (1) and (4) report OLS estimates; columns (2-3) and (5-6) report the IV analysis.

The IV analysis is performed again by using law DPR 81/2009. The law does not provide

[30] Older peers are children in higher grades (third, fourth, and fifth grades) attending the same multigrade class of second-grade students. Symmetrically, younger peers are only first graders attending the same multigrade class of second-grade students. Note that we do not have any test scores for first, third, or fourth graders as they do not take the INVALSI test. See Appendix A.1, part (b) for details about the process we use to identify such students.

specific rules guiding students' assignments to multigrade classes with younger or older peers; the decision rests solely with the school principal and it is based on the cohort size for each grade in each year. Therefore, this decision is usually driven by the number of students in adjacent grades. This provides us with a source of variation to construct a new set of instruments for attendance in specific (in terms of grade composition) multigrade classes. As an example, the probability of a second-grade student ending up in a multigrade class with younger peers depends on her cohort size (second grade) and the size of the younger adjacent cohort (first grade). We use this combined information to construct instruments. In detail, consider a case in which a second-grade student is enrolled in a school with fewer than 10 second graders. Now suppose that in the same school, the cohort size of first graders enrolled in the same SY is smaller than 15 students. In this specific case, the probability of the second graders ending up in a multigrade class with younger peers is considerably larger than in a case in which the cohort of first graders is larger than 15 students. With the same logic, we construct nine possible combinations suggested by the rules defined in DPR 81/2009, and we use indicators for these combinations as instruments for actual multigrading (and class size). For students in a multigrade class with younger peers, we consider the first grade as the adjacent grade, while for students sharing their class with older peers we consider the third grade as the adjacent grade.

First-stage estimates based on the new set of instruments are presented in Table A.6. As expected, the probability of ending up in a multigrade class is particularly high when both students' own and the adjacent grade cohort sizes are small (e.g. fewer than 10 students in their own grade and fewer than 15 students in the adjacent grade, or 10-14 students in their own grade and fewer than 10 in the adjacent grade).

In Table 11, both OLS and IV estimates suggest the same pattern in terms of the effect of grade composition in multigrade classes: the overall positive effect of multigrading is mainly driven by students sharing their (multigrade) class with more-mature peers. The OLS estimates pinpoint a statistically insignificant four percent of a standard deviation effect of multigrading with younger peers on child achievement. On the contrary, the effect is significantly larger (16 percent of standard deviation) for students sharing their class with older peers. The IV estimates confirm these findings. Sharing classes with less-mature students displays a positive and mostly statistically insignificant effect on child achievement. We find that the presence of older peers in the class determines a large and strongly significant positive effect on achievement (32–33 percent of a standard deviation).

According to this analysis, multigrading is particularly beneficial when a child shares the class with more-mature peers. This type of class is likely to inspire and foster a child's interactions with, and imitation of, her more-mature peers. At the same time, sharing the classroom with younger peers does not affect child development at this particular stage of individual growth.

6.2 Interpretation of the Multigrade Effect for Fifth Graders

In the baseline analysis, we find a statistically insignificant effect of multigrading for the sample of fifth graders.[31] However, the interpretation of such result is not straightforward. Students attending a multigrade class in their fifth (and last) year of primary school may be having their first experience with multigrading; alternatively, they may have been in a multigrade class for several years. These two possible scenarios would affect the interpretation of multigrading's zero impact. In the first case, the effect depends on attendance in a multigrade class during the last year of primary school.[32] In the second case, the multigrade variable would convey the cumulative effect of several years spent in a multigrade classroom. The analysis of these possible different channels is particularly challenging as many different student career paths are possible. Hence, the definition of the proper counterfactual for fifth graders is far from univocal. Indeed, it is possible that the control group, namely the group of students attending a single-grade class during their fifth year of primary school, also includes some students that attended multigrade classes during their school career.

[31] The result is similar to those for literacy in Checchi and De Paola (2018).

[32] It is important to note that this would imply fifth graders sharing the classroom with younger peers.

To deal with these potential concerns, we track back the grade composition of the classes attended by fifth graders in our sample.[33] Figure 3 shows the number of years in a multigrade class for students attending a multigrade class in their fifth year of primary school in the 2012/2013 SY. About half of students (49 percent) experienced multigrading in all five years of their primary school career. Three-fourth of them ended up in a multigrade in at least four out of the five years. This graphical evidence suggests the persistence of multigrading: fifth graders in a multigrade class are highly likely to have spent a considerable fraction of their primary school career in a multigrade class. This evidence supports the idea that the zero effect of multigrading for fifth graders is likely to convey the cumulative impact of multigrading along the entire primary school cycle.

To gain further confidence on this interpretation, we exploit the fact that the tracking of the entire primary school career path of fifth graders allows us to define alternative and more precise versions of the control and treatment groups. All the analyses performed consider the test score at the end of primary school (fifth grade) as the outcome of interest. We start with the analysis of the effect of having attended at least one year of multigrading during primary school. The control group consists of those students who never attended a multigrade class; the treatment group consists of those students who experienced at least one year of multigrading.[34] In this case, we obtain IV estimates with instruments constructed as the average of each student's cohort size over the five years of primary school. We then apply the standard rules defined by DPR 81/2009 to the average cohort size. Results in Table 12 display a zero effect of multigrading on student achievement at the end of primary school. Two important findings should be highlighted. First, this evidence reinforces the

[33]This procedure makes use of School Register data provided by MIUR for school years 2008/2009 through 2012/2013. We were able track back classes for 88,861 students, corresponding to 99 percent of our sample of fifth graders. In order to track back the school career of fifth graders, we assume that they did not change schools during their primary education. The share of students changing schools during the primary cycle is particularly low and it accounts for less than 3.5 percent of students in each grade. This value should be considered an upper bound, given that it includes changes within the same school when the school has different buildings in the same municipality.

[34]To correctly interpret the results, it is important to recall that persistence in multigrading is particularly high. This implies that in the majority of cases, attendance in at least one year of multigrading corresponds to several years in this kind of class.

interpretation that multigrading is a practice that is (at least) nondetrimental for child achievement. Second, that result is consistent with the mechanism shown for second graders and underlines the importance of sharing the class with older peers. Because multigrading is highly persistent, its cumulative effect is likely to be a combination of two opposite effects: a possible positive effect arising when the child belongs to the younger cohort in the classroom versus a possible negative effect when the child becomes part of the older cohort in the classroom.

We investigate the possible importance of sharing the classroom with more- or less-mature peers in Table 13. To this purpose, we analyze three possible different treatment groups: being part of a multigrade class exclusively in the last year of primary school (column 1), being part of a multigrade class exclusively in the fourth or fifth year (or both) of primary school (column 2), and being part of a multigrade class exclusively in the first or second year (or both) of primary school (column 3). We compare each of these groups with a control group consisting of those students who never attended a multigrade class during primary school. Unfortunately, persistence in multigrading limits the sample sizes for these groups and this limitation largely affects precision of IV estimates. For this reason, we only propose a suggestive correlation analysis based on OLS estimates. Although never statistically significant, correlations are negative when a student has been assigned to a multigrade class exclusively in the last years of primary school. These are the years in which the child is very likely to be part of the older cohort in the class. On the contrary, the multigrade coefficient turns positive for those students who experienced multigrading only at the very beginning of their career in primary school. These findings are consistent with the view that multigrading is likely to positively affect cognitive achievement of children sharing their class with more-mature peers.

The results for students that were in multigrade classes exclusively in the first years of primary school also allow us to infer some suggestive insights about the possible medium-run effects of multigrading. Indeed, in this framework, attendance in multigrade classes is assessed when the child is enrolled in her first or second grade, while test scores are measured when the child has reached the fifth grade. Here, the effect of multigrading amounts to around 11 percent of a standard deviation (statistically insignificant). This effect is remarkably similar to the OLS estimates of the effect of multigrading for second-grade students. Although this result seems suggestive of some persistence of the positive effect of multigrading for those students attending multigrade classes only as younger peers in the classroom, it only represents preliminary suggestive evidence. Further and more precise analysis on the possible medium-long run effects of attendance of multigrading classes is needed.

7 Conclusion

The development of cognitive and noncognitive skills in early childhood is recognized as a strong predictor of future success in academics as well as in life. For this reason, pedagogical practices that increase the abilities of young children represent powerful tools for improving individual well-being and reducing the chance of failure in the future.

Multigrading, the practice of placing children of different ages in the same classroom, is a common educational practice in both developing and developed countries. Although in developing countries it represents a widespread practice, in developed countries its use is generally confined to rural areas that are subject to population decline and where few children actually live. Nonetheless, over the last few years, multigrading has been adopted in several developed countries for reasons related to educational and pedagogical concerns. Supporters of this method emphasize its positive effects, including the benefits of a personalized approach to education, given that children of the same age can learn at different speeds. They also point to positive peer effects derived from younger children imitating their older peers' more responsible behavior. Although quite a common practice, the effects of multigrading on child achievement have rarely been carefully investigated because of possible endogeneity concerns.

In this paper, we analyze the effect of multigrading on children attending the second and fifth year of primary school in Italy. We do so by supplementing information on standardized test scores provided in the INVALSI data with information on schools and multigrade classes. To address endogeneity concerns and allow a proper causal inference, we exploit a national regulation by considering the number of students in the same cohort to determine whether a single or a multigrade class should be formed.

Our results suggest that multigrading positively affects cognitive achievement of second graders. The effect for fifth graders is close to zero. While the positive impact of multigrading for second-grade students mainly derives from students' sharing their class with more mature peers, the effect for fifth graders should be interpreted as the cumulative effect of multigrading over the entire cycle of primary schooling. Moreover, our results highlight that the positive multigrade effect for second graders is stronger for children from low socioeconomic backgrounds. These children, who are exposed to peers of different ages, receive important additional inputs that partially compensate for less-stimulating home environments.

This work suggests at least two main relevant policy implications. First, multigrading has potential beneficial effects for child development, especially when less-mature peers share their classroom with more-mature peers. Additionally, we do not find evidence of detrimental effects for those children sharing their classroom with less-mature peers. Moreover, evidence of heterogeneous impacts of multigrading with respect to children's socioeconomic background suggests that this practice gives more advantage to very young children coming from disadvantaged cultural and economic backgrounds. Therefore, according to our results, there is no supportive evidence against the practice of multigrading.

Second, multigrading is typical in schools in rural, sometimes remote, areas still common in Europe. These areas are generally characterized by low population density, deprivation, and abandonment by younger generations, which make their conditions poor. The adoption of multigrading may be the only way to leave these schools open, and schools are likely to be the only institution that has the potential to revitalize these areas.

References

Agostinelli, F. and G. Sorrenti (2018). Money vs. Time: Family Income, Maternal Labor Supply, and Child Development. HCEO Working Paper 2018-017.

Agostinelli, F. and M. Wiswall (2016). Estimating the Technology of Children's Skill Formation. NBER Working Paper No. 22442.

Angrist, J. D., E. Battistin, and D. Vuri (2017). In a Small Moment: Class Size and Moral Hazard in the Italian Mezzogiorno. *American Economic Journal: Applied Economics 9*(4), 216–49.

Angrist, J. D. and V. Lavy (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

Angrist, J. D., V. Lavy, and J. L.-L. A. Shany (2017). Maimonides Rule Redux. NBER Working Paper No. 23486.

Bertoni, M., G. Brunello, and L. Rocco (2013). When the Cat is Near, the Mice Won't Play: The Effect of External Examiners in Italian Schools. *Journal of Public Economics 104*, 65–77.

Bonesrønning, H. (2003). Class Size Effects on Student Achievement in Norway: Patterns and Explanations. *Southern Economic Journal 69*(4), 952–965.

Checchi, D. and M. De Paola (2018). The Effect of Multigrade Classes on Cognitive and Noncognitive Skills. Causal Evidence Exploiting Minimum Class Size Rules in Italy. *Economics of Education Review 67*, 235–253.

Cunha, F. and J. J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F. and J. J. Heckman (2010). Investing in Our Young People. In A. J. Reynolds, A. Rolnick, M. M. Englund, and J. Temple (Eds.), *Cost-Effective Early Childhood Programs in the First Decade: A Human Capital Integration*. Cambridge University Press.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Dahl, G. B. and L. Lochner (2012). The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *American Economic Review 102*(5), 1927–1956.

Diamond, R. and P. Persson (2016). The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests. NBER Working Paper No. 22207.

Dobbelsteen, S., J. Levin, and H. Oosterbeek (2002). The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition. *Oxford Bulletin of Economics and Statistics 64*(1), 17–38.

Ferrer-Esteban, G. (2012). Cheating to the Test in the Italian Standardized Assessment System: Rationale and Incentives. Working Paper Fondazione Giovanni Agnelli No. 50.

Gary-Bobo, R. J. and M. B. Mahjoub (2013). Estimation of Class-Size Effects, Using "Maimonides' Rule" and Other Instruments: The Case of French Junior High Schools. *Annals of Economics and Statistics* (111/112), 193–225.

Hargreaves, E., C. Montero, N. Chau, M. Sibli, and T. Thanh (2001). Multigrade Teaching in Peru, Sri Lanka and Vietnam: An Overview. *International Journal of Educational Development 21*(6), 499–520.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics 24*(3), 411–482.

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics 115*(4), 1239–1285.

ISTAT (2017). La Povertà in Italia - Anno 2016. ISTAT Statistiche Report.

Jacob, B. A. and S. D. Levitt (2003). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *The Quarterly Journal of Economics 118*(3), 843–877.

Leuven, E., H. Oosterbeek, and M. Rønning (2008). Quasi-Experimental Estimates of the Effect of Class Size on Achievement in Norway. *Scandinavian Journal of Economics 110*(4), 663–693.

Leuven, E. and M. Rønning (2016). Classroom Grade Composition and Pupil Achievement. *The Economic Journal 126*(593), 1164–1192.

Little, A. W. (2001). Multigrade Teaching: Towards an International Research and Policy Agenda. *International Journal of Educational Development 21*(6), 481–497.

Lucifora, C. and M. Tonello (2015). Cheating and Social Interactions. Evidence from a Randomized Experiment in a National Evaluation Program. *Journal of Economic Behavior & Organization 115*, 45–66.

Materiali UVAL (2014). A Strategy for Inner Areas in Italy: Definition, Objectives, Tools and Governance. In F. Barca, P. Casavola, and S. Lucatelli (Eds.), *Analisi e studi, Documenti, Metodi*. Materiali Uval.

Mulkeen, A. and C. Higgins (2009). Multigrade Teaching in Sub-Saharan Africa: Lessons from Uganda, Senegal, and the Gambia. World Bank Working Paper No. 173.

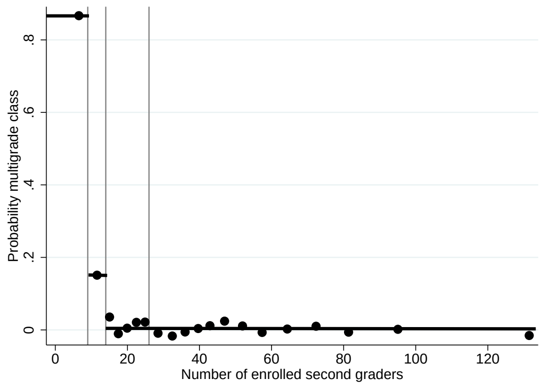
Todd, P. E. and K. I. Wolpin (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital 1*(1), 91–136.

UNESCO (2004). Educating Rural People: a Low Priority. Education Today, No. 9.

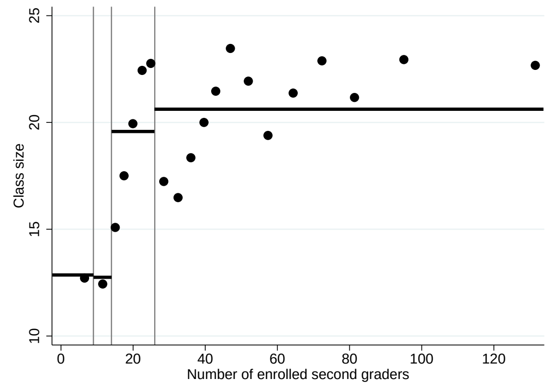
Figures and Tables

Figure 1: Number of Students Enrolled, Multigrading, and Class Size

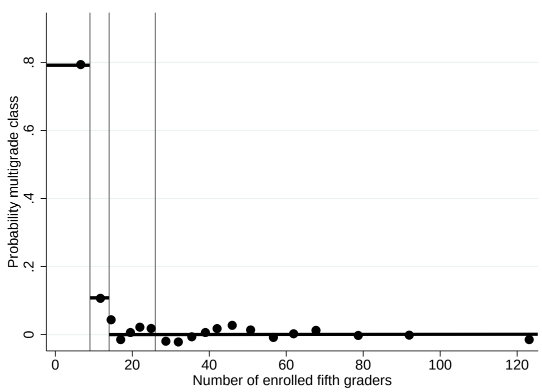
(a) 2^{nd} *Grade - Probability of multigrade*



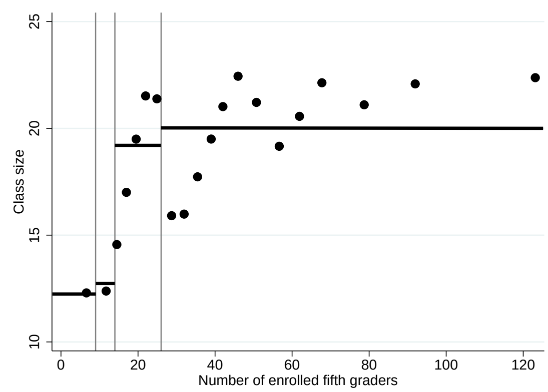
(b) 2^{nd} *Grade - Class size*



(c) 5^{th} *Grade - Probability of multigrade*



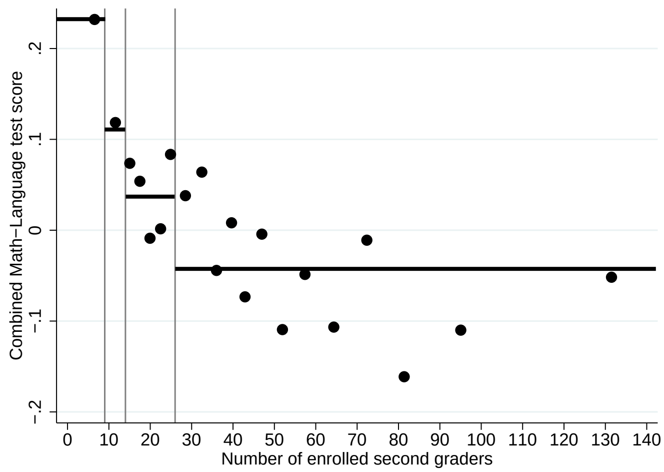
(d) 5^{th} *Grade - Class size*



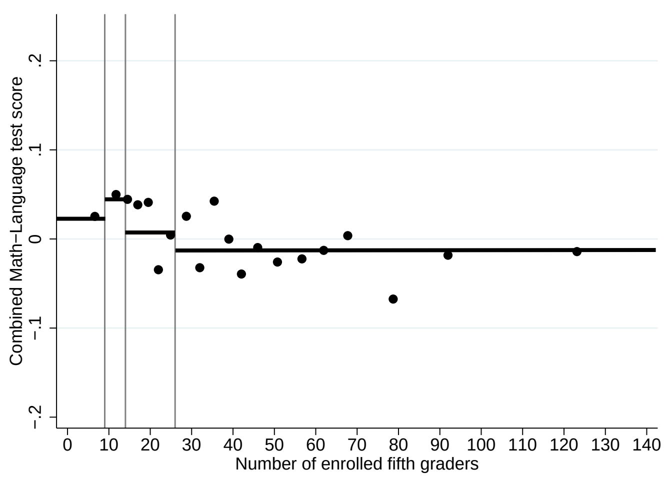
This figure shows the predicted individual probability of ending up in a multigrade class (Panels (a) and (c)) and average class size (Panels (b) and (d)) as functions of the number of (enrolled) students. Panels (a) and (b) refer to second-grade students. Panels (c) and (d) refer to fifth-grade students. Each bin represents the average y-axis value for each equal-size group (ventiles). Each solid line represents the average y-axis value for the four intervals in the number of students enrolled in a specific grade defined by DPR 81/2009: fewer than 10 students, 11–14 students, 15–26 students, at least 27 students. The predicted individual probability of attending a multigrade class (y-axis) is obtained through first-stage estimates in Table 4, column (2) for second-grade students. The predicted individual probability of attending a multigrade class (y-axis) is obtained through first-stage estimates in Table 5, column (2) for fifth-grade students. Refer to the text and to Tables 4 and 5 for further details about the empirical model underlying this figure.

Figure 2: Number of Students Enrolled and Performance on Standardized Tests

(a) 2^{nd} *Grade - Combined Math-Language test score*

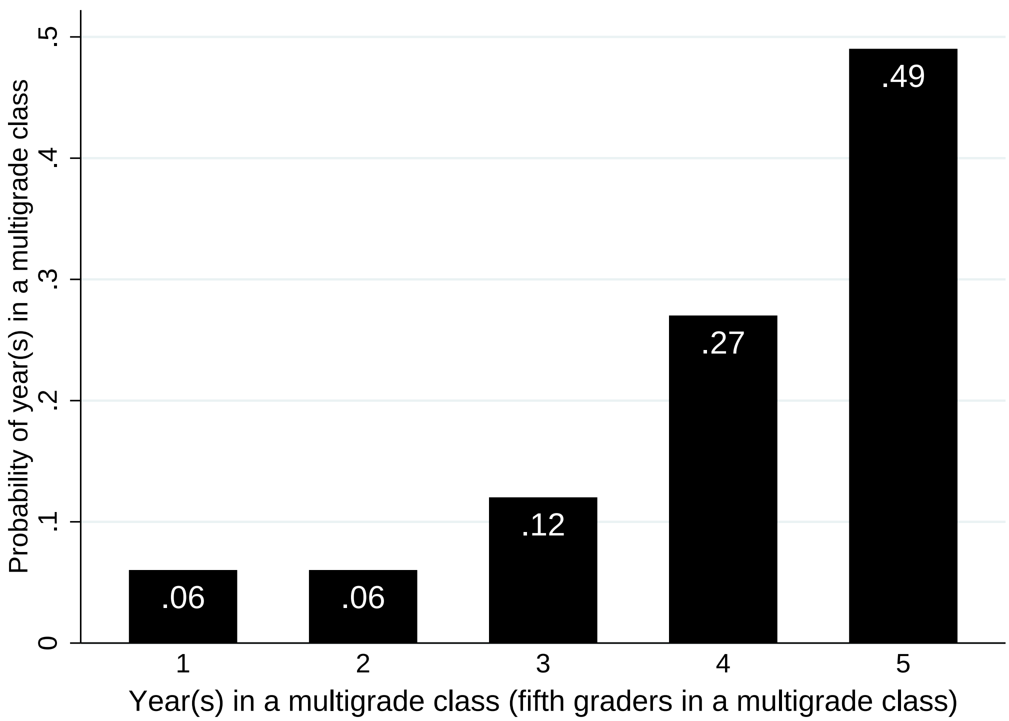


(b) 5^{th} *Grade - Combined Math-Language test score*



This figure shows the average performance on the standardized test (combined Math-Language) as a function of the number of (enrolled) students. The combined Math-Language test score has a mean of zero and standard deviation of one. Panel (a) refers to second-grade students. Panel (b) refers to fifth-grade students. Each bin represents the average y-axis value for each equal-size group (ventiles). Each solid line represents the average y-axis value for the four intervals in the number of students enrolled in a specific grade defined by DPR 81/2009: fewer than 10 students, 11–14 students, 15–26 students, at least 27 students.

Figure 3: The Persistence of Multigrading



This figure shows the number of years spent in a multigrade class for fifth graders in a multigrade class in the 2012/2013 SY. The figure is obtained by tracking back the entire school career of fifth-grade students in our sample.

Table 1: Summary Statistics

	Mean (1)	St.Dev. (2)	Mean (3)	St.Dev. (4)
Math	18.95	6.74	27.70	8.94
Language	24.84	6.73	62.82	12.48
Multigrade	0.06	0.23	0.05	0.22
Class size	19.45	4.09	18.86	4.00
Age	6.97	0.27	9.98	0.32
Female	0.49	0.50	0.50	0.50
Italian	0.88	0.32	0.89	0.31
Migrant 1st gen.	0.03	0.17	0.05	0.21
Migrant 2nd gen.	0.09	0.28	0.06	0.24
Father university	0.07	0.26	0.06	0.24
Father high school	0.75	0.44	0.76	0.43
Father other	0.18	0.39	0.18	0.38
Mother university	0.10	0.30	0.08	0.28
Mother high school	0.73	0.45	0.75	0.43
Mother other	0.17	0.38	0.17	0.37
Northwest	0.46	0.50	0.44	0.50
Northeast	0.17	0.38	0.17	0.37
Central area	0.11	0.32	0.11	0.32
South	0.18	0.38	0.20	0.40
Islands	0.08	0.27	0.08	0.27
Time distance (min.)	5.49	3.30	5.58	3.36
Population (2011)	4,673	3,097	4,609	3,057
Altitude	261	222	268	227
Sample	2^{nd} Grade		5^{th} Grade	
Observations	92,469		89,780	

Summary statistics for the samples analyzed in this work. Columns (1) and (2) refer to the sample of second-grade students; columns (3) and (4) refer to the sample of fifth-grade students.

Table 2: Balancing Test for Second-Grade Students

	Below Cutoff (BC) (1)	Above Cutoff (AC) (2)	BC-AC (3)	p-value (BC-AC) (4)
	First cutoff: 10 students			
Age	6.96	6.95	0.01	0.58
	(0.01)	(0.01)	(0.01)	
Female	0.48	0.49	-0.01	0.53
	(0.01)	(0.01)	(0.02)	
Italian	0.90	0.90	0.00	0.66
	(0.01)	(0.01)	(0.01)	
Father university	0.05	0.06	-0.01	0.82
	(0.01)	(0.01)	(0.01)	
Mother university	0.08	0.09	-0.01	0.36
	(0.01)	(0.01)	(0.01)	
Number of students	[8,9]	[10,11]		
	Second cutoff: 15 students			
Age	6.96	6.96	0.00	0.57
	(0.01)	(0.01)	(0.01)	
Female	0.47	0.49	-0.02	0.18
	(0.01)	(0.01)	(0.01)	
Italian	0.90	0.90	0.00	0.75
	(0.01)	(0.01)	(0.01)	
Father university	0.06	0.05	0.01	0.10
	(0.00)	(0.00)	(0.01)	
Mother university	0.10	0.10	0.00	0.98
	(0.01)	(0.01)	(0.01)	
Number of students	[13,14]	[15,16]		
	Third cutoff: 27 students			
Age	6.97	6.95	0.02	0.06
	(0.01)	(0.01)	(0.01)	
Female	0.49	0.48	0.01	0.46
	(0.01)	(0.01)	(0.01)	
Italian	0.90	0.88	0.02	0.06
	(0.01)	(0.01)	(0.01)	
Father university	0.06	0.07	-0.01	0.01
	(0.00)	(0.01)	(0.01)	
Mother university	0.09	0.10	-0.01	0.15
	(0.01)	(0.01)	(0.01)	
Number of students	[25,26]	[27,28]		

Comparison of students' characteristics just below (column 1) and just above (column 2) the three critical cutoffs of second-grade (enrolled) students identified by DPR 81/2009. The critical cutoffs are 10, 15, and 27 students. Interval widths around the cutoffs are defined by two students above/below (i.e. 8–9 students versus 10–11 students). The difference in means and the p-value for difference in means are reported in columns (3) and (4), respectively.

Table 3: Balancing Test for Fifth-Grade Students

	Below Cutoff (BC) (1)	Above Cutoff (AC) (2)	BC-AC (3)	p-value (BC-AC) (4)
	First cutoff: 10 students			
Age	9.94	9.95	-0.01	0.20
	(0.01)	(0.01)	(0.01)	
Female	0.51	0.49	0.02	0.19
	(0.01)	(0.01)	(0.02)	
Italian	0.92	0.90	0.02	0.01
	(0.01)	(0.01)	(0.01)	
Father university	0.05	0.05	0.00	0.98
	(0.00)	(0.00)	(0.01)	
Mother university	0.07	0.08	-0.01	0.39
	(0.01)	(0.01)	(0.01)	
Number of students	[8,9]	[10,11]		
	Second cutoff: 15 students			
Age	9.97	9.96	0.01	0.30
	(0.01)	(0.01)	(0.01)	
Female	0.50	0.50	0.00	0.94
	(0.01)	(0.01)	(0.01)	
Italian	0.91	0.91	0.00	0.77
	(0.00)	(0.00)	(0.01)	
Father university	0.05	0.06	-0.01	0.23
	(0.00)	(0.00)	(0.01)	
Mother university	0.07	0.08	-0.01	0.31
	(0.00)	(0.00)	(0.01)	
Number of students	[13,14]	[15,16]		
	Third cutoff: 27 students			
Age	9.96	9.97	-0.01	0.52
	(0.01)	(0.01)	(0.01)	
Female	0.49	0.51	-0.02	0.16
	(0.01)	(0.01)	(0.01)	
Italian	0.90	0.89	0.01	0.55
	(0.01)	(0.01)	(0.01)	
Father university	0.06	0.05	0.01	0.59
	(0.00)	(0.00)	(0.01)	
Mother university	0.08	0.08	0.00	0.39
	(0.01)	(0.01)	(0.01)	
Number of students	[25,26]	[27,28]		

Comparison of students' characteristics just below (column 1) and just above (column 2) the three critical cutoffs of fifth-grade (enrolled) students identified by DPR 81/2009. The critical cutoffs are 10, 15, and 27 students. Interval widths around the cutoffs are defined by two students above/below (i.e. 8–9 students versus 10–11 students). The difference in means and the p-value for difference in means are reported in columns (3) and (4), respectively.

Table 4: First-Stage Estimates for Second-Grade Students

	Model (1) Multigrade OLS (1)	Model (2) Multigrade OLS (2)	Model (3) Multigrade OLS (3)	Model (3) Class size OLS (4)
$2^{nd}Graders < 10$	0.86***	0.90***	0.86***	-5.19***
	(0.01)	(0.01)	(0.01)	(0.20)
$10 \leq 2^{nd}Graders < 15$	0.15***	0.19***	0.15***	-5.44***
	(0.01)	(0.02)	(0.01)	(0.18)
$15 \leq 2^{nd}Graders < 27$	-0.00	-0.01***	-0.00	0.97***
	(0.00)	(0.00)	(0.00)	(0.17)
Class size		0.01***		
		(0.00)		
Father university	-0.00	-0.00	-0.00	-0.03
	(0.00)	(0.00)	(0.00)	(0.05)
Mother university	-0.00	-0.00	-0.00	-0.01
	(0.00)	(0.00)	(0.00)	(0.04)
SW Chi-sq. (UId)	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00
SW F (WId)	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00
KP (WId)	> 100	> 100	> 100	> 100
F-stat ($2^{nd}Gr. < 10$)	> 100	> 100	> 100	> 100
F-stat ($10 \leq 2^{nd}Gr. < 15$)	> 100	> 100	> 100	> 100
F-stat ($15 \leq 2^{nd}Gr. < 27$)	0.14	10.23	0.14	31.28
Instrumented variable(s)	Multigrade	Multigrade	Multigrade, Class size	
Sample	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade
Observations	92,469	92,469	92,469	92,469

First-stage estimates. Dependent variable: Being enrolled in a multigrade class (columns 1,2, and 3), class size (column 4). The reference category for the number of second-grade students is $2^{nd}Graders \geq 27$. The reference category for father's and mother's education is completed high school. Model (1) does not include class size as a control variable. Model (2) includes class size as a control variable. Model (3) treats both multigrade and class size as endogenous variables. All models include controls for child's gender, age, nationality, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5: First-Stage Estimates for Fifth-Grade Students

	Model (1) Multigrade OLS (1)	Model (2) Multigrade OLS (2)	Model (3) Multigrade OLS (3)	Model (3) Class size OLS (4)
$5^{th}Graders < 10$	0.79***	0.83***	0.79***	-5.06***
	(0.02)	(0.01)	(0.02)	(0.20)
$10 \leq 5^{th}Graders < 15$	0.10***	0.15***	0.10***	-4.97***
	(0.01)	(0.01)	(0.01)	(0.16)
$15 \leq 5^{th}Graders < 27$	-0.00	-0.01***	-0.00	1.09***
	(0.00)	(0.00)	(0.00)	(0.18)
Class size		0.01***		
		(0.00)		
Father university	-0.00*	-0.00*	-0.00*	-0.01
	(0.00)	(0.00)	(0.00)	(0.05)
Mother university	-0.00	-0.00	-0.00	-0.06
	(0.00)	(0.00)	(0.00)	(0.04)
SW Chi-sq. (UId)	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00
SW F (WId)	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00
KP (WId)	> 100	> 100	> 100	> 100
F-stat ($5^{th}Gr. < 10$)	> 100	> 100	> 100	> 100
F-stat ($10 \leq 5^{th}Gr. < 15$)	86.66	> 100	86.66	> 100
F-stat ($15 \leq 5^{th}Gr. < 27$)	2.58	25.06	2.58	38.64
Instrumented variable(s)	Multigrade	Multigrade	Multigrade, Class size	
Sample	5^{th} Grade	5^{th} Grade	5^{th} Grade	5^{th} Grade
Observations	89,780	89,780	89,780	89,780

First-stage estimates. Dependent variable: Being enrolled in a multigrade class (columns 1,2, and 3), class size (column 4). The reference category for the number of fifth-grade students is $5^{th}Graders \geq 27$. The reference category for father's and mother's education is completed high school. Model (1) does not include class size as a control variable. Model (2) includes class size as a control variable. Model (3) treats both multigrade and class size as endogenous variables. All models include controls for child's gender, age, nationality, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Multigrading and Child Achievement

	Combined Math-Language							
	OLS (1)	IV (2)	IV (3)	IV (4)	OLS (5)	IV (6)	IV (7)	IV (8)
Multigrade (M)	0.09***	0.24***	0.16***	0.19***	-0.02	0.05	0.01	-0.01
	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.05)
Class size (CS)	-0.01***		-0.01***	-0.01	-0.01**		-0.01**	-0.01
	(0.00)		(0.00)	(0.01)	(0.00)		(0.00)	(0.00)
$E(CS\|M=0)-$ $E(CS\|M=1)$	5.94	5.94	5.94	5.94	5.30	5.30	5.30	5.30
Instrumented variable(s)		Multigrade	Multigrade	Multigrade, Class size		Multigrade	Multigrade	Multigrade, Class size
Sample	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	5^{th} Grade	5^{th} Grade	5^{th} Grade	5^{th} Grade
Observations	92,469	92,469	92,469	92,469	89,780	89,780	89,780	89,780

OLS and IV estimates of the effect of multigrading on a child's test score. Dependent variable: Combined Math-Language test score. Models (1) to (4) refer to second-grade students. Models (5) to (8) refer to fifth-grade students. Models (2) and (6) do not include class size as a control variable. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. $E(CS|M=0) - E(CS|M=1)$ represents the difference in the average class size (number of students) between single-grade classes ($M = 0$) and multigrade classes ($M = 1$). Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 7: Different Sets of Instruments for Multigrading and Class Size

	Combined Math-Language	
	IV (1)	IV (2)
Multigrade	0.19***	-0.01
	(0.05)	(0.05)
Class size	-0.01	-0.01
	(0.01)	(0.00)
	First stage Multigrade	
$Students < 10$	0.86***	0.79***
	(0.01)	(0.02)
$10 \leq Students < 15$	0.15***	0.10***
	(0.01)	(0.01)
F-stat ($Students < 10$)	> 100	> 100
F-stat ($10 \leq Students < 15$)	> 100	87.46
	First stage Class size	
$Students < 15$	-5.34***	-5.00***
	(0.18)	(0.17)
$15 \leq Students < 27$	0.97***	1.09***
	(0.17)	(0.18)
F-stat ($Students < 15$)	> 100	> 100
F-stat ($15 \leq Students < 27$)	31.11	38.82
Instrumented variables	Multigrade, Class size	Multigrade, Class size
Sample	2^{nd} Grade	5^{th} Grade
Observations	92,469	89,780

IV estimates of the effect of multigrading on a child's test score. Dependent variable: Combined Math-Language test score. Model (1) refers to second-grade students. Model (2) refers to fifth-grade students. *Students* refers to the the number of (enrolled) second-grade (model 1) and fifth-grade students (model 2). The reference category in the first stage for multigrade is $Students \geq 15$. The reference category in the first stage for class size is $Students \geq 27$. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Distance to the Closest School as Instrument for Parents' Preferences

	Combined Math-Language			
	IV (1)	IV (2)	IV (3)	IV (4)
Multigrade	0.16***	0.19***	0.00	-0.02
	(0.04)	(0.05)	(0.04)	(0.05)
Class size	-0.01***	-0.01	-0.01**	-0.01
	(0.00)	(0.01)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Sample	2^{nd} Grade	2^{nd} Grade	5^{th} Grade	5^{th} Grade
Observations	92,469	92,469	89,780	89,780

IV estimates of the effect of multigrading on a child's test score. Dependent variable: Combined Math-Language test score. Models (1) and (2) refer to second-grade students. Models (3) and (4) refer to fifth-grade students. Road distance (in minutes) to the closest alternative school is used as an additional instrument for being enrolled in a multigrade class and class size. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality and geographical macro-area. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 9: Multigrading and Child Achievement: The Case of Northern Regions

	Combined Math-Language					
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Multigrade	0.09**	0.11**	0.15**	-0.05	-0.05	-0.05
	(0.04)	(0.05)	(0.06)	(0.03)	(0.04)	(0.06)
Class size	-0.01***	-0.01***	-0.00	-0.01**	-0.01**	-0.01
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Instrumented variable(s)		Multigrade	Multigrade, Class size		Multigrade	Multigrade, Class size
Regions	Only North	Only North	Only North	Only North	Only North	Only North
Sample	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	5^{th} Grade	5^{th} Grade	5^{th} Grade
Observations	58,329	58,329	58,329	54,402	54,402	54,402

OLS and IV estimates of the effect of multigrading on a child's test score. Dependent variable: Combined Math-Language test score. Models (1) to (3) refer to second-grade students. Models (4) to (6) refer to fifth-grade students. The analysis is only based on northern Italian regions. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 10: Heterogeneous Effects of Multigrading on Child Achievement

	Combined Math-Language			
	Child's gender		Parental education	
	IV (1)	IV (2)	IV (3)	IV (4)
	Female	Male	No one with university	One with university
Multigrade	0.22***	0.17***	0.20***	0.09
	(0.06)	(0.06)	(0.05)	(0.07)
Class size	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Instrumented variables	Multigrade, Class size	Multigrade, Class size	Multigrade, Class size	Multigrade, Class size
Sample	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade
Observations	45,224	47,245	79,856	12,613
	Female	Male	No one with university	One with university
Multigrade	0.03	-0.05	-0.02	0.06
	(0.05)	(0.06)	(0.05)	(0.09)
Class size	-0.01	-0.01	-0.01*	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Instrumented variables	Multigrade, Class size	Multigrade, Class size	Multigrade, Class size	Multigrade, Class size
Sample	5^{th} Grade	5^{th} Grade	5^{th} Grade	5^{th} Grade
Observations	44,875	44,905	79,262	10,518

Heterogeneous analysis by child's gender (models 1 and 2) and parental background (models 3 and 4). Dependent variable: Combined Math-Language test score. The upper panel refers to second-grade students. The lower panel refers to fifth-grade students. All models include controls for child's gender (except models 1 and 2), age, nationality, father's and mother's educational level (except models 3 and 4), and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 11: Grade Composition and Child Achievement

	Combined Math-Language					
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Multigrade	0.04	0.11*	0.10	0.16***	0.33***	0.32***
	(0.04)	(0.06)	(0.07)	(0.05)	(0.09)	(0.10)
Class size	-0.01***	-0.01***	-0.01**	-0.01***	-0.01***	-0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Instrumented variable(s)		Multigrade	Multigrade, Class size		Multigrade	Multigrade, Class size
In multigrade with	Younger peers	Younger peers	Younger peers	Older peers	Older peers	Older peers
Sample	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade
Observations	89,461	89,461	89,461	88,428	88,428	88,428

Analysis of the effect of multigrading according to class composition in terms of grades. Dependent variable: Combined Math-Language test score. Younger peers means that only first-grade students attend the same multigrade class of second-grade students. Older peers means that children of higher grades (third, fourth, and fifth grades) attend the same multigrade class of second-grade students. Instruments used in models (2) and (3) are indicators for different intervals in the joint distribution of first-grade and second-grade (enrolled) students. Instruments used in models (5) and (6) are indicators for different intervals in the joint distribution of second-grade and third-grade (enrolled) students. Refer to the text for further details and to Table A.6 for first-stage estimates. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 12: The Effect of Spending at Least One Year in a Multigrade Class

	Combined Math-Language		
	OLS (1)	IV (2)	IV (3)
At least one year in multigrade	-0.01	0.01	-0.04
	(0.03)	(0.03)	(0.05)
Class size	-0.01**	-0.01**	-0.01**
	(0.00)	(0.00)	(0.01)
Instrumented variable(s)		Multigrade	Multigrade, Class size
Sample	5^{th} Grade	5^{th} Grade	5^{th} Grade
Observations	88,861	88,861	88,861

OLS and IV estimates of the effect of spending at least one year in multigrading on a child's test score. Dependent variable: Combined Math-Language test score. The analysis is obtained by tracking back the entire primary school career of fifth-grade students. Instruments are obtained by averaging each student's cohort size over the five years of primary school. The rules established by DPR 81/2009 are then applied to the average cohort size as in the rest of the paper to obtain the instruments. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 13: The Effect of Attending a Multigrade Class in Different Grades

	Combined Math-Language		
	OLS (1)	OLS (2)	OLS (3)
Multigrade (M)	-0.05	-0.04	0.11
	(0.11)	(0.07)	(0.08)
Class size	-0.00**	-0.01**	-0.00**
	(0.00)	(0.00)	(0.00)
Sample	5^{th} Grade	5^{th} Grade	5^{th} Grade
Treatment group (TG)	M in 5^{th} G	M in 4^{th} or 5^{th} G	M in 1^{st} or 2^{nd} G
Control group	Never in M	Never in M	Never in M
Share TG	< 0.01	0.01	< 0.01
Observations	82,681	83,220	82,772

OLS estimates of the effect of attending a multigrade class in different primary school grades on a child's test score. Dependent variable: Combined Math-Language test score. The analysis is obtained by tracking back the entire primary school career of fifth-grade students. The control group always consists of fifth-grade students that have always attended a single-grade class during primary school. The treatment groups are: students in a multigrade class only at their fifth year of primary school (column 1), students in a multigrade class only at their fourth or fifth (or both) year of primary school (column 2), and students in a multigrade class only at their first or second (or both) year of primary school. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, and road distance in time to the closest alternative school. M denotes multigrade classes. Share TG represents the share of students in a (specific, see definition above) multigrade class with respect to the total sample size. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Appendix: Additional Material

A.1 Data Construction Process

In this Appendix, we describe the process we used to: a) identify students attending multigrade classes, b) identify the grade composition of multigrade classes, and c) track the career path of fifth-grade students.

a) Students in multigrade classes

As mentioned in the paper, the INVALSI data do not contain information concerning which type of class a student attended, so students enrolled in a multigrade class cannot be directly identified. To obtain this information, we merged three administrative archives. The first data set (the INVALSI data from now on) contains information about children's performance on the INVALSI test in school year 2012/2013. For each student, the test score in both mathematics and language as well as background information such as gender, age, nationality, attendance at preparatory schools, and parents' education and profession are available. Neither school names nor school characteristics and location are available in this data set. However, each individual record also includes a class and a school code, as well as geographical and demographic information about the municipality where the student's school is located. This piece of information is fundamental for our matching procedure and includes: i) the province where the school is located, ii) the population (in the 2001 and the 2011 census) of the municipality, iii) the size (in square km) of the municipality, and iv) the altitude of the municipality where the school is located.

A second administrative data set (School Register data from now on) provided by the Italian Ministry of Education contains detailed information about the characteristics of each Italian primary school in school year 2012/2013 as well as the five previous years. All the Italian regions are covered in this data with the exception of Valle d'Aosta and Trentino Alto Adige. The School Register includes information such as school name, municipality, number of students (total and in each grade), number of classes (total and in each grade), and number of multigrade classes. Based on this information, we analyzed all of the possible combinations of grade composition at the school level to identify different types of schools. For example, if a school shows a positive number of second-grade students, but no second-grade single-grade classes and at least one multigrade class, we can assume that second-grade students attend a multigrade class. We ended up with: i) schools where second-grade students attend a multigrade class; ii) schools where second-grade students attend one second-grade single-grade class; iii) schools where more than one second-grade single-grade class is offered; and iv) schools with no second-grade students. Note that we found no evidence of primary schools with both single and multigrade classes for the same grade.

Unfortunately, the INVALSI data and the School Register data cannot be matched directly. In fact, the first data set only identifies each primary school with an anonymous code. The only way to overcome this problem is to identify (at least) the names of the municipalities where the schools included in the INVALSI data set are located. Once identified, it would be possible to match the data set with the School Register, with municipality as the matching variable.

The Municipality Register data set provided by ISTAT is the last piece of information needed to complete the data construction process. The Municipality Register contains geographical and demographic information for each Italian municipality. This information (province, population in the 2001 and 2011 census, size and altitude of the municipality) is the same as that contained in the INVALSI data, therefore making the merger of the INVALSI data set with the Municipality register data set possible. We use geographical and demographic information as key identifying variables in the matching process to obtain the INVALSI+ISTAT data.

Last, we match the INVALSI+ISTAT data with the School Register data based on municipality names. Unfortunately, with this last matching, we are able to uniquely identify only schools located in municipalities hosting no more than one school. We repeated the same procedure to obtain the data for fifth-grade students.

b) Grade composition of multigrade classes

As mentioned in the paper, no data identify the grade composition of multigrade classes. We use the data built in the previous paragraph and apply a wide set of rules to identify grade composition of multigrade classes. These rules are based on the information originally included in the School Register. For example, we define the following *Rule 1* to identify a multigrade class whose students are first and second graders only (therefore second graders are the older peers in the multigrade class). According to *Rule 1*, the school has:

- a) one multigrade class;
- b) no first- and second-grade single classes;
- c) first- and second-grade students;
- d) third-, fourth-, and fifth-grade single classes;
- e) third-, fourth-, and fifth-grade students.

We consider about 40 such rules to enumerate all the possible combinations of students of different grades and to describe the classes in our data.

c) School career path of fifth-grade students in SY 2012/2013

To interpret the effect of multigrading on fifth-graders, we tracked back their school career path. We repeated the procedure described in point (b) for fourth-grade students in SY 2011/2012, third-grade students in SY 2010/2011, second-grade students in SY 2009/2010, and first-grade students in SY 2008/2009. This procedure allowed us to determine how many years (and which years) the students attending fifth-grade in SY 2012/2013 spent in a multigrade or in a single-grade class (assuming that they did not change schools during their primary education).

A.2 Additional Figures and Tables

Table A.1: Reduced-Form Estimates

	Combined Math-Language			
	OLS (1)	OLS (2)	OLS (3)	OLS (4)
$Students < 10$	0.17***	0.23***	0.02	0.05
	(0.04)	(0.04)	(0.04)	(0.03)
$10 \leq Students < 15$	0.05	0.11***	0.03	0.06*
	(0.04)	(0.03)	(0.03)	(0.03)
$15 \leq Students < 27$	0.05*	0.04	0.02	0.01
	(0.03)	(0.03)	(0.02)	(0.02)
Class size	-0.01***		-0.00*	
	(0.00)		(0.00)	
Sample	2^{nd} Grade	2^{nd} Grade	5^{th} Grade	5^{th} Grade
Observations	92,469	92,469	89,780	89,780

Reduced-form estimates. Dependent variable: Combined Math-Language test score. Models (1) and (2) refer to second-grade students. Models (3) and (4) refer to fifth-grade students. The world *Students* refers the the number of (enrolled) second-grade (models 1 and 2) and fifth-grade students (models 3 and 4). The reference category for the number of students is $Students \geq 27$. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.2: First-Stage Estimates With Distance to the Closest School as Instrument

	Model (1)	Model (2)		Model (3)	Model (4)	
	Multigrade	Multigrade	Class size	Multigrade	Multigrade	Class size
	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
$2^{nd} Graders < 10$	0.90***	0.86***	-5.19***	0.83***	0.79***	-5.06***
	(0.01)	(0.01)	(0.20)	(0.01)	(0.02)	(0.20)
$10 \leq 2^{nd} Graders < 15$	0.19***	0.15***	-5.44***	0.15***	0.10***	-4.97***
	(0.02)	(0.01)	(0.18)	(0.01)	(0.01)	(0.16)
$15 \leq 2^{nd} Graders < 27$	-0.01***	-0.00	0.97***	-0.01***	-0.00	1.09***
	(0.00)	(0.00)	(0.17)	(0.00)	(0.00)	(0.18)
Time distance	0.00**	0.00**	-0.02	0.00	0.00	-0.01
	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.02)
Class size	0.01***			0.01***		
	(0.00)			(0.00)		
SW Chi-sq. (UId)	> 100	> 100	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00	0.00	0.00
SW F (WId)	> 100	> 100	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00	0.00	0.00
KP (WId)	> 100	> 100	> 100	> 100	> 100	> 100
F-stat ($2^{nd} Gr. < 10$)	> 100	> 100	> 100	> 100	> 100	> 100
F-stat ($10 \leq 2^{nd} Gr. < 15$)	> 100	> 100	> 100	> 100	86.66	> 100
F-stat ($15 \leq 2^{nd} Gr. < 27$)	10.23	0.14	31.28	25.06	2.58	38.64
F-stat (Time distance)	5.24	4.30	0.92	0.67	0.51	0.15
Instrumented variable(s)	Multigrade	Multigrade, Class size		Multigrade	Multigrade, Class size	
Sample	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	5^{th} Grade	5^{th} Grade	5^{th} Grade
Observations	92,469	92,469	92,469	89,780	89,780	89,780

First-stage estimates. Dependent variable: Being enrolled in a multigrade class (columns 1, 2, 4, and 5), class size (columns 3 and 6). The reference category for the number of second-grade students is $2^{nd} Graders \geq 27$. The reference category for the number of fifth-grade students is $5^{th} Graders \geq 27$. Models (1) and (3) include class size as control variable. Models (2) and (4) treat both multigrade and class size as endogenous variables. Road distance (in minutes) to the closest alternative school is used as an additional instrument for being enrolled in a multigrade class and class size. All models include controls for child's gender, age, nationality, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.3: Multigrading and Cohort Size

	Combined Math-Language							
	IV (1)	IV (2)	IV (3)	IV (4)	IV (5)	IV (6)	IV (7)	IV (8)
Multigrade	0.15***	0.19***	0.19***	0.21***	0.02	-0.01	0.00	0.00
	(0.04)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Class size	-0.01***	-0.01	-0.01***	-0.01	-0.00**	-0.01	-0.01***	-0.01*
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Cohort size	$\leq$ 60 Stud.	$\leq$ 60 Stud.	$\leq$ 30 Stud.	$\leq$ 30 Stud.	$\leq$ 60 Stud.	$\leq$ 60 Stud.	$\leq$ 30 Stud.	$\leq$ 30 Stud.
Sample	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	5^{th} Grade	5^{th} Grade	5^{th} Grade	5^{th} Grade
Observations	69,629	69,629	37,399	37,399	68,891	68,891	36,980	36,980

IV estimates of the effect of multigrading on a child's test score. Dependent variable: Combined Math-Language test score. Models (1) to (4) refer to second-grade students. Models (5) to (8) refer to fifth-grade students. Models (1) and (2) only include schools with at most 60 second-grade (enrolled) students. Models (3) and (4) only include schools with at most 30 second-grade (enrolled) students. Models (5) and (6) only include schools with at most 60 fifth-grade (enrolled) students. Models (7) and (8) only include schools with at most 30 fifth-grade (enrolled) students. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.4: Math and Language Test Scores

	IV (1)	IV (2)	IV (3)	IV (4)
	Panel (a): Math			
Multigrade	0.17***	0.20***	0.02	-0.00
	(0.04)	(0.05)	(0.04)	(0.05)
Class size	-0.01***	-0.01	-0.01**	-0.01*
	(0.00)	(0.01)	(0.00)	(0.01)
	Panel (b): Language			
Multigrade	0.12***	0.14***	0.00	-0.01
	(0.04)	(0.05)	(0.03)	(0.04)
Class size	-0.01***	-0.01	-0.00**	-0.01
	(0.00)	(0.01)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Sample	2^{nd} Grade	2^{nd} Grade	5^{th} Grade	5^{th} Grade
Observations	92,469	92,469	89,780	89,780

IV estimates of the effect of multigrading on a child's test score. Dependent variable: Math test score (Panel (a)), Language test score (Panel (b)). Models (1) and (2) refer to second-grade students. Models (3) and (4) refer to fifth-grade students. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.5: Sensitivity Analysis

	Combined Math-Language			
	IV (1)	IV (2)	IV (3)	IV (4)
	Panel (a): Parents' missing information			
Multigrade	0.16***	0.17***	-0.02	-0.05
	(0.04)	(0.05)	(0.04)	(0.05)
Class size	-0.01***	-0.01	-0.01***	-0.01**
	(0.00)	(0.01)	(0.00)	(0.01)
	Panel (b): Regional (NUTS 2) FE			
Multigrade	0.18***	0.20***	0.02	-0.01
	(0.04)	(0.05)	(0.04)	(0.05)
Class size	-0.01***	-0.01	-0.00**	-0.01*
	(0.00)	(0.01)	(0.00)	(0.00)
	Panel (c): Provincial (NUTS 3) FE			
Multigrade	0.20***	0.21***	0.02	-0.01
	(0.04)	(0.05)	(0.04)	(0.05)
Class size	-0.01***	-0.01*	-0.01**	-0.01*
	(0.00)	(0.01)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Sample	2^{nd} Grade	2^{nd} Grade	5^{th} Grade	5^{th} Grade

Sensitivity analysis for baseline estimates. Dependent variable: Combined Math-Language test score. Models (1) and (2) refer to second-grade students. Models (3) and (4) refer to fifth-grade students. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area (except Panels (b) and (c)), and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

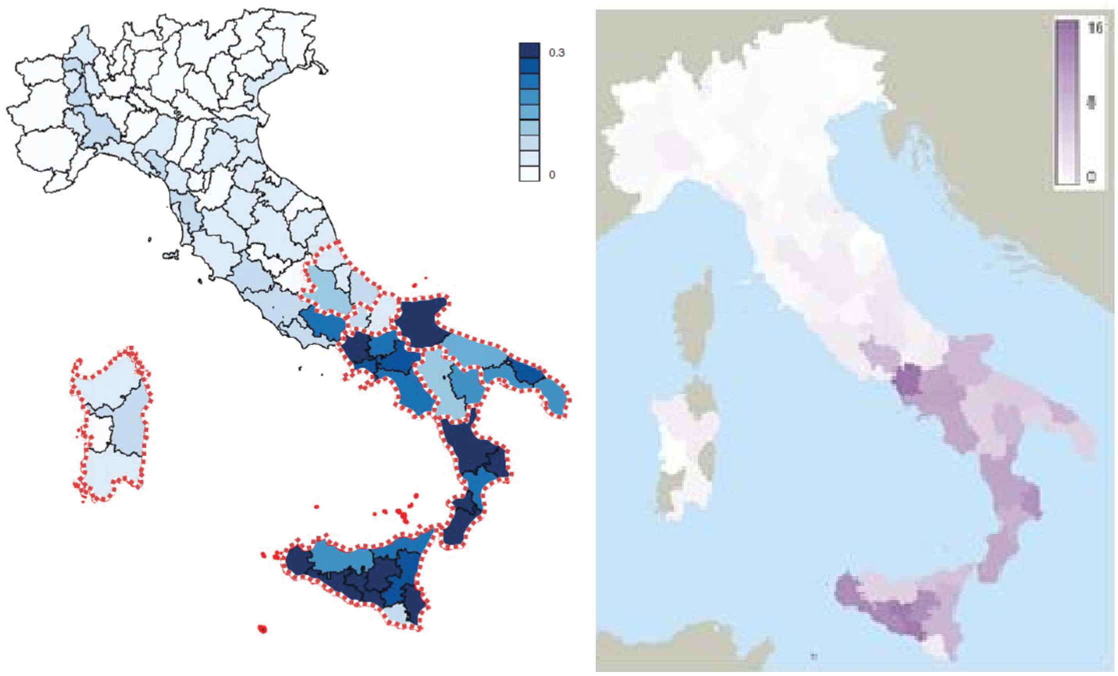
Table A.6: First-Stage Estimates of Grade Composition

	Model (1) Multigrade OLS (1)	Model (2) Multigrade OLS (2)	Class size OLS (3)	Model (3) Multigrade OLS (4)	Model (4) Multigrade OLS (5)	Class size OLS (6)
$2^{nd}G < 10$ & $AdjG < 15$	0.79***	0.77***	-5.65***	0.67***	0.65***	-6.39***
	(0.02)	(0.03)	(0.24)	(0.03)	(0.04)	(0.26)
$2^{nd}G < 10$ & $15 \leq AdjG < 27$	0.07**	0.03	-9.59***	0.25***	0.22***	-8.50***
	(0.03)	(0.03)	(0.37)	(0.07)	(0.07)	(0.50)
$2^{nd}G < 10$ & $AdjG \geq 27$	(Not observed in the sample)			0.04***	0.00	-9.95***
				(0.01)	(0.00)	(0.21)
$10 \leq 2^{nd}G < 15$ & $AdjG < 10$	0.39***	0.37***	-5.13***	0.20***	0.18***	-5.34***
	(0.04)	(0.04)	(0.23)	(0.04)	(0.04)	(0.32)
$10 \leq 2^{nd}G < 15$ & $10 \leq AdjG < 15$	0.07***	0.05***	-5.84***	0.05***	0.02***	-5.96***
	(0.02)	(0.01)	(0.20)	(0.01)	(0.01)	(0.19)
$10 \leq 2^{nd}G < 15$ & $AdjG \geq 15$	0.02***	0.00	-5.83***	0.06***	0.04***	-5.56***
	(0.00)	(0.00)	(0.17)	(0.01)	(0.01)	(0.19)
$15 \leq 2^{nd}G < 27$ & $AdjG < 10$	0.07*	0.07*	-0.56*	0.01***	0.00	-1.09***
	(0.04)	(0.04)	(0.33)	(0.00)	(0.00)	(0.33)
$15 \leq 2^{nd}G < 27$ & $AdjG \geq 10$	-0.00**	0.00	1.02***	-0.00***	0.00	1.05***
	(0.00)	(0.00)	(0.17)	(0.00)	(0.00)	(0.17)
Class size	0.00***			0.00***		
	(0.00)			(0.00)		
SW Chi-sq. (UId)	> 100	> 100	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00	0.00	0.00
SW F (WId)	> 100	> 100	> 100	62.86	> 100	> 100
P-value	0.00	0.00	0.00	0.00	0.00	0.00
KP (WId)	> 100	> 100	> 100	62.86	> 100	> 100
Instrumented variable(s)	Multigrade	Multigrade, Class size		Multigrade	Multigrade, Class size	
In multigrade with	Younger peers	Younger peers	Younger peers	Older peers	Older peers	Older peers
Sample	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade	2^{nd} Grade
Observations	89,461	89,461	89,461	88,428	88,428	88,428

First-stage estimates of grade composition. Dependent variable: Being enrolled in a multigrade class with younger (columns 1, 2) or older peers (columns 4, 5), and class size (columns 3, 6). Younger peers means that only first-grade students attend the same multigrade class of second-grade students. Older peers means that children of higher grades (third, fourth, and fifth grades) attend the same multigrade class of second-grade students. $2^{nd}G$ stands for number of second-grade (enrolled) students. $AdjG$ stands for number of (enrolled) students in the adjacent grade. The adjacent grade is the first grade in models (1) and (2), and the third grade in models (3) and (4). The reference category for the set of instruments is $2^{nd}G \geq 27$. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Figure A.1: Territorial Distribution of Cheating

(a) *Score manipulation rates by province* (b) *Cheating*



This figure shows the geographical distribution of cheating according to two different sources. Panel (a) refers to the work by Angrist et al. (2017) and is based on: (i) implausible score levels, (ii) the within-class average and standard deviation of test scores, (iii) the number of missing items, and (iv) a Herfindahl index of the share of students with similar response patterns. Panel (b) refers to the work by Ferrer-Esteban (2012) and is based on the analysis of the sequence of identical answers at the class level as a signal for possible cheating.